

## Chapter 12

# Proficiency-Based Progression Simulation Training: A To-Do List for Medicine

*There is no excuse for the surgeon to learn on the patient*  
William H. Mayo M.D. (pp. 1378, 1927)

Change has been the order of the day in medicine, but particularly in disciplines such as surgery. Surgery has changed the way it treats patients with interventions becoming less invasive but also becoming more difficult to learn and to practice. Sometimes these changes were patient driven. One of these changes, minimally invasive surgery (MIS) was introduced on a wave of enthusiasm in the early 1990s (Centres 1991). It was a disruptive technology and had unforeseen and wide-reaching implications and ramifications for the entire practice of medicine. In the original description of this phenomenon, the authors argued that “disruptive innovations can hurt successful, well managed companies that are responsive to their customers and have excellent research and development. These companies tend to ignore the markets most susceptible to disruptive innovations (Bower and Christensen 1995).” That is how MIS took hold of the field of surgery, i.e., patient demand. The complications that were associated with the practice of this new type of surgery became very public and pointed to a skills deficit in the operating surgeon. It is unfortunate for surgery that these developments occurred around about the same time as high-profile medical errors cases were being investigated (e.g., The Bristol Case (Senate of Surgery 1998)) in the UK and the “To Err is Human” Report (Kohn et al. 2000) in the USA. We believe that both of these developments had a profound influence on medicine for the better. The introduction of MIS forced the surgical community to investigate why this type of surgery was more difficult to learn than the traditional open approach, and as a result surgery in particular and medicine in general had to closely examine how they prepared doctors to treat patients. The high-profile error cases forced the medical community to confront an uncomfortable truth which is that some patients are made sicker or die as a direct result of the care they are given by their doctor. While this was not a new phenomenon the patients were being told about it on the media. Worse still was that in some cases, the public were told that the medical community knew about “it” and did nothing until the issue had been made public. The hemorrhage of public confidence from medicine as a result of these incidents cannot be underestimated.

As a result of the investigations into medical errors, it became clear that a high proportion of them occurred in surgery. Regenbogen et al. (2007) have suggested that between one half and two thirds of hospital adverse events are attributable to surgery and surgical care. Also, the sorts of errors that occur in surgical care tend to be different from those that occur on medical services, making many of the studies of medication errors in hospital not easily generalizable to surgical care. The big difference is that most surgical errors occur in the operating room and most are technical in nature. Technical errors are errors in which some aspect of the surgery is not done properly and concern manual skills and errors of surgical judgment or knowledge. Surgery is unique among medical specialties in that while doing operations, surgeons are constantly making decisions in real time and acting on them. These sorts of errors can occur at any phase of surgical care and have been attributed to low hospital volume, breakdown in communications, systems failures, fatigue, lack of experience in trainees and many other causes. The results from the Regenbogen et al. (2007) study are not unique. Similar results have been reported in Belgium using a similar type of research methodology. Somville et al. (2010) retrospectively reviewed surgical malpractice claims from 3,202 malpractice liability cases, in which patients alleged error, between 1996 and 2006. They identified surgical errors that resulted in patient injury in 427 study claims. The results showed that 63% of these cases involved a significant or major error injury and 6% involved death. In most cases (48%), errors occurred in intraoperative care, 15% in preoperative care and 37% in postoperative care. The leading factors which were associated with errors were inexperience/lack of technical competence (57%) and communication breakdown (42%). Furthermore, cases involving technical errors were more likely to occur during elective surgery. These findings were not available at the time surgery and medicine were conducting root and branch analysis of how they practice medicine; however, they serve as reinforcement that the analyses was appropriate.

## Training Efficiently

Whether as a result of these medical errors or as an evolution of common sense in medical training, the number of hours which junior doctors are required to work have been reduced dramatically. This did not happen in one country, but in almost every country with a well-developed medical training system. Neither of us can recall going to a conference during the last decade and NOT hearing a senior surgeon bemoaning the reduction in training hours for junior surgeons. The same is true in other disciplines in medicine. No amount of complaining will change the situation regarding training hours. What is rarely discussed by leaders in medicine is the inefficiency of the current training system. In the USA, it takes 5–7 years to train a surgeon, assuming they undertake a Fellowship in their specialty. In the UK and Ireland, it takes between 11 and 13 years to train. The question should be asked: is the performance of surgeons in the USA who finish after 5 or 7 years so much

inferior to the performance of surgeons who finish in the UK and Ireland after 11–13 years? This may not be a politically polite question.

The simple fact of the matter is that surgery and medicine are training doctors for twenty-first century medicine using a nineteenth century training paradigm. Halsted developed and implemented his apprenticeship model in the late nineteenth century because there was nothing else available in the USA that was as systematic and presumably effective. He did what he could with the resources he had available to him. At the start of the twenty-first century, we are duty bound to build on Halsted's legacy. We know considerably more about how human beings learn, how they acquire knowledge and skills, the limits of their sensory and perceptual system, and how all of these human factors can be facilitated and augmented to better achieve "education and training." By constructing an apprenticeship-based curriculum for surgical training, Halsted was configuring and organizing the information that the trainee acquired which in turn facilitated them in becoming a safe surgeon. What we have proposed here in this book simply builds on that methodology. In the past, medicine was learned from books, lectures, tutorials, and practicals. It was also learned from repeated practice on real patients. The methodology that we are proposing here really does not differ significantly from what has been done in the past in terms of content. However, where it does significantly depart from what has been done in the past is how that content is delivered. We have argued that content alone does not make an education and training program effective. What makes education and training effective and efficient is how the content is delivered and how the delivery is configured. Human beings are not simply passive information processors; they are not simply vessels that we can pour knowledge and skills into (mores the pity). This means that when we teach trainees, we cannot assume that they have learned the material or understand it, nor can we assume that they can do something that we trained them to do (never mind do it to a certain standard).

Human beings are more likely to remember information that has been organized for them and has been sequenced in a logical and meaningful order. Furthermore, we cannot assume that they have learned the material; we must check. Likewise, skill acquisition should be organized in a sequential and sensible fashion where basic skills are acquired before more complex skills and performance must be assessed. The trainee must know how they are performing and the trainer must know how a trainee is progressing. The trainee will learn fastest and most efficiently if they have formative feedback during their training. Furthermore, for training to be effective, trainees cannot simply engage in repeated practice; they must engage in deliberate practice. Deliberate practice differs from repeated practice in terms of how training is configured but more importantly or the formative and summative feedback that the trainee is given.

## **Proficiency-Based Progression**

This information is not new but what is new is how it is applied to the acquisition and practice of procedural skills such as surgery and to those that would suggest that we are just spoon-feeding the trainees, we would point out that what we are

advocating is simply good educational and training practices that are well grounded in quantitative research. If anything, our proposals place a greater emphasis on the effort made by the trainee. Ericsson et al. (1993) have shown that performance excellence is not something that individuals are born with rather it is something that has been acquired over 10 years of deliberate practice. Many surgical trainees will find this an uncomfortable truth. What we have proposed here is that deliberate practice should be used for the effective and efficient acquisition of skills and knowledge. However, this process cannot be continued ad *infinitum* by educational and training institutions. That is more the responsibility of the trainee, and the regulatory agencies have been particularly good at ensuring continuing professional development as an integral and non-negotiable part of medicine. We have suggested that training should continue until the trainee has reached a performance criterion level. Furthermore, that performance criterion level should be quantified on the basis of real-world surgical/medical skills. Unfortunately, there continues to be too much ambiguity and debate about precisely what constitutes “competency.” To circumvent these issues, we have objectively defined and quantitatively assessed proficiency. Dreyfus et al. (1986) have proposed that proficient skills are those that have been developed to a stage beyond competent skills. This means that if skills are demonstrated to be proficient, by default, they must be competent. To quantify the performance level of proficiency, we have used the performance of experienced practicing surgeons. There can be little doubt that the vast majority of these individuals’ performance is at least competent. Using this approach, we have been able to establish a quantitative goal for the trainee based on the real skills of real practicing surgeons. It also means that the benchmark that has been established is fair, objective and transparent. Furthermore, it is a sufficiently flexible approach to training to allow the gifted trainee to progress through the training cycle quicker than those trainees who take longer to reach the level of proficiency. Moreover, it does not discriminate against the trainees who acquire their skills at a slower rate. The ethos of training is that once the proficiency level has been demonstrated (consistently), that part of training is completed. The other advantage with proficiency-based progression training is that it ensures that ALL individuals in the training program have successfully demonstrated the required skill level. This is not the case with the traditional training approach. Unfortunately, in the traditional training approach, the same amount of time in training is presumed to fit all when it is obvious that this is not the case.

A proficiency-based progression training paradigm places the onus on the trainers to provide the facilities and the learning resources for the trainee to acquire the skills and knowledge to learn their craft. However, it places the onus on the trainee to unambiguously demonstrate that they have reached the prescribed level of performance. This approach to training is far removed from the “spoon-feeding” approach that some individuals might so caricature. This is a relatively new approach to training, and few assumptions are made about the knowledge and skill level of the graduating trainee. Rather, they must demonstrate that they have the knowledge and skills before graduating; otherwise they do not progress. The development of metrics for the assessment process on which proficiency is established will be new to most of

medicine. However, it is a well-established and validated protocol in the behavioral sciences (Kazdin 1994; Martin et al. 1993). Furthermore, it is relatively straightforward, and once users have experienced the entire process a couple of times, they will develop a comfortable familiarity with it. It is a process that Halsted would probably have been comfortable with because it pays attention to detail. In fact, the effectiveness of the training and assessment system relies on reliably capturing performance detail. The thesis behind the system is that proficient surgeons are good at what they do because of their attention to small but apparently inconsequential details of task performance which they probably perform automatically and unthinkingly. However, it is the attention of the surgeon to these details that makes their performance proficient or better. For example, it probably does not make that much difference when suturing a wound closed whether or not all of the knots are aligned on one side of the wound, whether or not they are spaced equally apart and the suture tails are approximately equal (not too short, not too long), etc. However, it is attention to these types of detail that probably typify the approach of the operator to other and less inconsequential aspects of the procedure.

What was demonstrated in the past is that if a trainee has been trained to the level of proficiency which has been based on the performance scores of experienced and practicing surgeons (in that particular task or procedure), those trainees outperform their peers who have gone through a traditional curricular training program (Ahlberg et al. 2007; Seymour et al. 2002; Van Sickle et al. 2008b). These studies have been prospective randomized and blinded in their assessment of the proficiency-based progression training paradigm. Although the subject numbers in each of the studies were small, the differences between the traditionally and the proficiency-based progression trained surgeons were large. Some surgeons may claim that the number of subjects in the studies were too small from which to generalize the results. In response to this, we would point out that science is about the unambiguous establishment of cause-and-effect relationships. These studies have unambiguously demonstrated in a prospective, randomized, and blinded fashion that proficiency-based progression trainees perform better.

## **Metric Validation**

We have no illusions that there will be critics of this approach to training, and in the best traditions of the scientific enterprise we will be the first to celebrate the verification of an alternative strategy with the same scientific rigor that has been applied to proficiency-based progression. One of the cornerstones of proficiency-based progression training is the performance metrics. These will be developed from rigorous task analyses by experienced groups of surgeons proficient at performing the surgical task or operation in question. The performance characteristics that they identify during the task analysis will be explicitly operationally defined in a way that they are refutable. This is a crucial aspect of an objective, transparent and fair assessment system. We have been critical of assessment strategies which are less explicit,

e.g., OSATS (Martin et al. 1997). Although we are sympathetic to their goals, attempting to score the performance characteristics of procedural medicine such as surgery on a Likert scale is more difficult than it should be, using OSATS. Trying to establish high inter-rater reliability using a Likert scale scoring system is almost impossible or, at least, will take more time to accomplish than most consultant surgeon assessors are prepared to give. It is much easier to establish high levels of inter-rater reliability with a checklist scoring system. However, the checklist that has been constructed for the assessment of performance on any task or procedure must be comprehensive and incisive. Furthermore, it needs to be valid. The metrics that have been identified as part of the task analysis should be shown to distinguish between the performance of experts and novices or at least experienced practitioners and novices. If metric-based performance does not distinguish between these groups, the metrics are flawed, and probably important aspects of the performance of the procedure have not been well characterized. However, we have not encountered a set of metrics that have been developed using the methodology that we have described that did not distinguish between experts and novices (with one exception). If a surgical task is so simple that a brief explanation and one demonstration is sufficient to transfer the skills and knowledge to a trainee, construct validity (i.e., being able to show a difference between the performance of experts and novices) will be difficult to demonstrate (indeed, we would suggest pointless).

Surgery and other procedural-based disciplines in medicine must move away from ambiguous definitions of performance characteristics. They are difficult to measure and have the tendency to allow bias and possibly even unfair practices to creep into the assessment system. There is some evidence that the new assessment systems that are being introduced into the training programs in the UK are becoming more explicit about what they assess. The DOPS system uses a Likert-type scale for the assessment of performance; however, it is only used for formative assessments (Chap. 7). For high stakes assessment, such as PBAs, a checklist scoring system is used (Chap. 8). However, attempting to reliably assess performance characteristics that have been defined as, “optimum” (without definition), “adequate,” “sound,” and “purposeful” leaves too much room for individual interpretation and will almost certainly impact on inter-rater reliability levels.

In Chap. 7 we examined the issue of inter-rater reliability levels in great detail. That was because these are the metric units of performance on which trainees within a training program will be passed or failed. In our opinion, the least that the person being assessed can expect is that the examiners agreed on at least 80% of their assessment scores (as the performance characteristics have been defined). It does not mean that the assessors agree 80% of the time for the entire class that is being assessed; it does not mean that the correlation between the two examiners scores is  $r \geq 0.8$ , nor does it mean that the alpha coefficient between the two raters is  $\geq 0.8$ . However, that is what some researchers are reporting in validation studies (Bann et al. 2003; Khan et al. 2007; Larsen et al. 2006) in some of the highest impact journals in surgery and medicine. Inter-rater reliability means the percentage of agreement between the two examiners on the individual who is being assessed. Anything less rigorous than this approach to validation may lead to successful litigation claims by trainees whose

training progress has been halted because they failed to demonstrate proficiency using metrics that had been validated using a validation process other than 80% agreement between assessors. Proficiency-based progression training ensures the quality of performance of the trainee. However, it also makes the system that they are being assessed by much more transparent than it has been in the past. Furthermore, these assessments are not called high stakes by coincidence; these assessments determine whether the trainee progresses in their training. Some trainees who fail to progress will almost certainly seek legal redress as they will have already invested many years in education and training. Anything less than transparently rigorous validation of all levels of proficiency-based progression training programs, and in particular the metric-based assessment units, will lead to successful legal challenges. Ironically, it is easier to get the process right than it is to do it wrong!

## Proficiency Refined

The skill acquisition framework that we have proposed here derives from the model proposed by Dreyfus and Dreyfus (Dreyfus et al. 1986). Although the Dreyfus and Dreyfus model proposes a conceptual framework, it does not offer nor advocate a measurement strategy. The quantification strategy that we dovetailed with this model comes from the behavioral sciences and has been used for more than half a century. We are satisfied that they complement each other well; however, we do have some philosophical questions that have practical implications about the characterization and implementation of proficiency-based training. Proficiency as characterized by us is the performance of experienced surgeons; these individuals are experienced in performing the task or the surgical procedure which we wish to set a level of proficiency. They, preferably are, not the leading surgeons in the world at performing the task or procedure and likewise they are not at the opposite end of that scale. Rather, their performance lies somewhere around the middle of that performance spectrum. Metrics that are developed from the analysis of the task or procedure should be capable of characterizing the performance of these individuals to the extent that it can reliably distinguish between their performance and that of novices or less experienced operators. This may seem imprecise and that is because it is. We developed this strategy to avoid the alternative which is the development and application of standardized operating procedures. The methodology is robust enough to ensure that it is fairly representative of the vast majority of operating surgeons who perform the procedure or task; however, it also sets a high enough standard so that trainees who reach that level perform significantly better intraoperatively than trainees who go through the traditional training program. Furthermore, benchmarks established on the performance of these experienced surgeons appear to be reachable by the vast majority of surgical trainees who persist in deliberate practice training sessions.

The first time a proficiency-based progression training strategy (based on the methodology that we have described here) was used was in the original VR to OR

study conducted at Yale University in the USA (Seymour et al. 2002). In that study, virtual reality training subjects trained on the simulator in a 1-h session until they reached the performance criteria level (or level of proficiency) with both hands, on two consecutive trials. The reasoning was:

1. That the surgical task that they were to perform (i.e., dissection of the gallbladder from the liver bed using electrocautery) was a bimanual task and therefore they had to be equally skilled with both hands.
2. They had to reach the level of proficiency on two consecutive trials, because they could demonstrate proficiency once potentially by accident, but not twice in a row.
3. Proficiency was quantitatively defined on the basis of five attending surgeons' performance on the training task.
4. Furthermore, it was the mean performance of the surgeons that constituted the performance criteria levels for errors and economy of diathermy.

One of the problems that we have with the characterization of proficiency as described here is that for the trainees to demonstrate proficiency, must, on average, perform better than 50% of the surgeons on whom proficiency was quantified. Furthermore, why does proficiency have to be the mean of the performance of the experienced operators? why could it not be the mean plus one standard deviation, or indeed the median? Also, why has proficiency to be demonstrated on two consecutive trials; why not more than two? These are questions that need to be quantitatively addressed probably sooner than later. An alternative strategy would be to investigate the receiver operating characteristic (ROC) of proficiency development and the clinical implications of adopting different training strategies. ROC analysis provides tools to select possible optimal models and to discard suboptimal ones independently from (and prior to specifying) the cost context or the class distribution. ROC analysis is related in a direct and natural way to cost/benefit analysis of diagnostic decision making. The ROC curve was first developed by electrical engineers and radar engineers during World War II for detecting enemy objects in battle fields, also known as the signal detection theory and was soon introduced in psychology to account for perceptual detection of signals (Swets 1996). Whatever strategy is eventually decided upon, it will be a difficult balancing act to fulfill. The level of proficiency must be conservative enough to ensure that it confers a uniform and high standard of intraoperative performance that optimizes patient safety. The standard must not be set so high that trainees find it very difficult, if not impossible to reach. The way that proficiency is currently construed appears to work fairly well, but we believe that it can be improved further.

## **Proficient Experts?**

One of the problems that relates to the quantitative definition of proficiency is the much wider issue of objective assessment of technical performance in surgery. Much of the methodology that we have discussed in this book is about the objective and fair assessment of performance and how this might be approached.



This approach was then validated and the validated metric units were used to establish performance benchmarks. These benchmarks were based on the performance of experienced operators. The assumption being that these experienced operators were “good” at what they did. What do we do if they are not? This is not a hypothetical situation. One of the first studies to report the performance of some surgeons who are performing significantly worse than their peers was by Gallagher et al. (2003c). They found that some other surgeons who participated in the study could not complete any part of the relatively simple box trainer and virtual reality laparoscopic tasks. Furthermore, some of those who were able to complete the tasks were performing more than 20 standard deviations from the mean. The data were checked and rechecked; the relationship between the operative experience of the surgeon and their performance was checked, as was the reliability of the simulator. All of these potential explanations were rejected as reasonable explanations for the performance of this small group of surgeons. These surgeons’ performances were always *more* than two standard deviations worse than their peers and frequently worse than the trainees to whom we were comparing them for the establishment of construct validity! The alternative explanation was that they simply performed badly on the tasks on the day that they were assessed and that this probably bore no resemblance to their intraoperative clinical performance.

As the years have passed and more experience in the objective assessment of the surgical skills has been accrued, this explanation also seems unlikely. For a small minority of surgeons that we have encountered, there appears to be no correlation between their objectively assessed performance and their self-reports of their own intraoperative performance. We have not systematically nor aggressively pursued a scientific answer to this question even though we suspect we know the answers. However, from personal experience, we believe that a strong correlation does exist between objectively assessed performance and intraoperative performance. Informally acquired information on some operators (e.g., surgeons, interventional cardiologists, etc.) seems to corroborate the suspicion that individuals who do not perform well in the skills laboratory also perform poorly intraoperatively. If these were trainees, there really would not be a problem. The problem arises from the fact that these individuals are consultant or attending surgeons. These are the very individuals whom we wish to benchmark so that we can use their performance as a training goal for their juniors. Take for example, a consultant surgeon who when objectively assessed is performing five standard deviations worse than their peers. The ethos of the proficiency-based progression training program is that proficiency should be established on the basis of experienced operators’ performance, and therefore, their performance measures should be included in the proficiency definition. After all, these individuals are experienced operators. How should these individuals be dealt with?

We are not sure how to deal with them. In general, the surgical community are aware that these individuals exist, but in the past it was extremely difficult to quantify their performance other than in terms of bad outcomes and their outcomes were not “significantly” worse than some of their peers. That situation has changed and we can now reliably and validly assess intraoperative performance which simple

logic dictates has to be related to intraoperative performance. Some in the surgical community might argue that our intraoperative performance characterization (e.g., metric-based assessment) does not really capture the performance of a surgeon and the hypothesized relationship between objectively assessed intraoperative performance and outcomes has never been established. However, the intraoperative performance metrics that we use to assess performance have been identified by a group of experienced operators who have identified characteristics that they believe distinguish between optimal and suboptimal performance. Van Sickle et al. (2008a) found that when they compared the objectively assessed intraoperative performance of attending surgeons to surgical residents on an intracorporeal suturing task, the intraoperative metrics reliably distinguished between the groups of surgeons. Furthermore, these types of detailed task performance metrics constitute the same types of parameters that the aviation industry uses in their analysis of near misses. The logic that the aviation industry uses is that each near miss is an accident waiting to happen. As pointed out previously, the performance units that we use in the objective assessment of performance may be better construed as “events” which are best defined by their outcomes but that each event set the occasion for a potential bad outcome to occur. These are what Reason (2000) refers to as the latent conditions in the chain of error causation. It should also be recalled that Reason was very clear that latent conditions are much easier to deal with than active failures. In essence, the technically poor performing surgeon is the latent condition that sets the occasion for active failures. Also, as discussed in Chaps. 4, 8, and 10, surgeons who struggle with relatively straightforward skills-based scenarios will not be able to cope with intraoperative clinical situations that are more demanding. In one sense, it is not their fault as they simply do not have the cognitive attentional resources to deal with the situation. However, who should recognize and act appropriately with this as a potential latent error situation: the surgeon? the hospital? their profession? A previous head of department once said that if he ignored some problems long enough, they just went away. We strongly suspect that this one would not and will in fact probably get worse as more and more evidence accrues linking bad outcome to the intraoperative performance of the operator. Also, this is not just a problem for surgery but for all of procedural-based medicine. Surgery just happens to be grasping the nettle first. We are fully aware that bad things happen to good surgeons and are very sympathetic to this view. Surgeons and other interventionalists have a very difficult and complex job to do. Unlike many other medical disciplines, they have to perform well technically while at the same time having to make difficult intraoperative decisions 'on-the-fly'. When many surgeons see a bad outcome happening to one of their peers, they think “there but for the grace of god go I.” The surgeons with whom these infrequent events occur are not the surgeons we are alluding to.

Our approach to individuals who perform badly on the objective assessment is simply to exclude them from the proficiency definition process and take the matter no further. After all, their performance does not accurately reflect the vast majority of their peers' performance. Furthermore, the rule of thumb that we use in the exclusion is performance that is more than 1.96 standard deviations away from the mean (in a

negative direction). It could be argued that performance in a positive direction creates as much of a problem; but to date, we have not found this to be the case. Not everyone is happy with this approach, least of all the person who has been excluded from the proficiency definition. However, there is little else that can be done at this stage. This is not simply a matter for the surgical community to resolve. We have made the same observations in other procedural-based disciplines in medicine. The scientific issue that begs to be resolved is the unambiguous establishment of a relationship and the strength of that relationship between objectively assessed intraoperative performance and clinical outcomes. This question is answerable. The study would need to be very large and conducted independently in the countries around the world who carry major responsibilities for training large numbers of procedural-based specialists. It should also be noted that the vast majority of operating surgeons have absolutely nothing to fear from this process. It will quantitatively confirm what we already know and that is that the majority of operating surgeons perform similar to their peers. A small number will be outstanding performers and a very small minority will demonstrate considerable skills deficits.

## **Regional, National, and International Levels of Proficiency**

In the USA, the American Boards of Surgery and Internal Medicine, etc., are responsible for the examination and licensure of surgeons and physicians across the entire country. Currently their examination system consists mainly of knowledge and decision-making assessments. However, with wider acceptance of the validity of technical skills assessment, it offers the opportunity to standardize assessment of this aspect of surgical performance across the USA. Furthermore, these assessment and credentialing boards are well known for the rigor with which they apply to the assessment process. This assessment process could be used as a liberal inclusion process rather than a conservative exclusion process. However, the outcome would almost certainly mean that individuals whose technical performance may best be characterized as “outliers” would almost disappear. In the USA, re-credentialing is a non-negotiable part of practicing as a doctor. This process would also ensure less performance variability across the country. The data could also be used to establish where on the performance distribution surgical graduates from other countries lay. The process could even facilitate the credentialing of international surgical graduates who wished to work in the USA. Although no equivalent credentialing system exists in the UK and Ireland, there are urgent plans to implement a similar system. One of the problems that the UK and Irish system have is that surgical graduates from outside the jurisdiction are entitled to apply for training positions and jobs. However, there is little or no way of objectively establishing how good, bad, or indifferent is the applicant’s performance. A valid and reliable system for the assessment of technical skills would considerably simplify answering that question. This would ensure a much fairer approach to the applicant and an even fairer treatment of the patient.

This approach to credentialing has other, possibly less attractive ramifications for procedural specialties like surgery. We have possibly seen a glimpse of the future in the FDA decision on carotid artery stenting with an embolic protection device. In the rollout of this relatively new approach to treatment for carotid artery disease, vascular surgeons who normally treated this condition found themselves in competition with interventional cardiologists, interventional radiologists, interventional neuroradiologists, and neurosurgeons. The decision of the FDA and The Centers for Medicare and Medicaid Services (or CMS) was that all interested medical specialties who could demonstrate proficiency in performing the procedure could claim reimbursement (Gallagher and Cates 2004a). This was made possible because part of the FDA decision included an acceptance that proficiency could be achieved in part by training on a high-fidelity virtual reality simulation. Furthermore, rather than simply relying on procedural numbers, proficiency demonstration on the simulator could be underpinned with metric-based performance characterization. Although the FDA decision related to the marketing and sale of the device, the impact radiates outward to medical practice as no physician of any procedure specialty could use the device in the absence of other skills associated with making appropriate interventional judgments about the patient's care. The physician may be proficient in the use of the device, capable of deploying the device in the correct fashion, but the physician may still not be allowed to perform the procedure. To ensure safe care of patients, an operating physician requires patient-specific knowledge of the anatomy, pathophysiology, treatment effects, and robust knowledge of the overall clinical status of the patient. Simulator training may be necessary for proficiency to be demonstrated, but simulator training alone is not sufficient for a physician to be certified as competent to perform interventional care (Dawson 2006).

Dawson (2006) also argues that simulator-based training is not a replacement for clinical experience. We tend to disagree with him on this point. We agree with him that simulation will not entirely replace clinical experience. However, it will supplant a large part of it particularly in the early stage of the learning curve where it is very difficult to justify basic procedural training on a sick patient. The full impact and ramifications of the FDA decision have not been fully realized yet. However, the FDA decision has levelled the playing field in terms of which medical specialty can perform interventional procedures. We believe that this decision will impact on who can be credentialed to perform other procedures such as colonoscopy, natural orifice total endoscopic surgery and a wide range of new percutaneous endovascular procedures. The FDA decision means that large governmental organizations now know that they do not have to take an individual physician's or medical specialties' word about their capability to perform a given procedure safely. They can now insist on quantitative evidence to demonstrate this fact. We are not entirely sure where this development is going to lead but we feel certain that it will have profound implications for the practice of safe interventional medicine. The FDA decision may have no legal implications outside the USA but precedents are difficult to ignore when grappling with similar issues in similar circumstances.

## What is the Relationship Between Proficiency and Competency?

In the Dreyfus et al. (1986) model of skill acquisition, they describe proficiency as a more advanced stage of skill acquisition than competency. Their proposal is a useful heuristic in trying to conceptualize the process of learning skills. However, their proposal contributes very little to the operational definition and measurement of the different levels of skills development that they outline. What they propose for the different levels of skills development are nothing more than descriptive indicators which are really not much better than the descriptions of competency outlined by the Accreditation Council for Medical Education (Beall 1999) in the USA and the General Medical Council (1993) in the UK. The clinical trials conducted on proficiency-based progression training have avoided the term “competency”-based progression because of the lack of an unambiguous and agreed-upon definition of what is competence. Ironically, the operational definition of competence is purely a matter of words and agreement within the medical profession itself. The difference between the concept of proficiency that we propose here and that has been operationalized in previous clinical trials (Ahlberg et al. 2007; Seymour et al. 2002; Van Sickle et al. 2008b) is that there is a general consensus among physicians and surgeons that doctors currently in practice are at least competent, probably proficient, and some are expert. The other difference is that proficiency has been quantitatively defined based on the performance of doctors whom most people agree are competent and/or proficient. Hence, the definition is parsimonious, i.e., proficiency is what proficient doctors do. This means that by default, proficiency has already been quantified for some tasks and surgical procedures. Furthermore, this methodology has been validated both in terms of metric validation and clinical validation. Would this approach solve the impasse on the issue of competence? We suspect not.

The issues that medicine has about competency are not to do with measurement they are more to do with agreeing on a definition. Once a benchmark has been set for the measurement of competence, the logical conclusion of this process means that some individuals will be measured as “not competent.” There is considerable trepidation among physicians and surgeons about this eventuality even though as stated earlier that the majority of practitioners have absolutely nothing to fear. Our concern is that at some point, medicine may be forced to quantitatively define competence at a time and over an issue that is not of medicines choosing. At some point, someone, possibly a legislator, possibly a failed trainee, possibly the very wealthy parents of a failed trainee, is going to ask, “When exactly is someone deemed competent or conversely when are they deemed incompetent?” An individual who failed to progress in the competency-based training system in the USA or in the UK must have failed to demonstrate one or more specific competencies. The concept of competency, if it is to be at all meaningful, must be verifiable and falsifiable (Popper 1979). That is probably one of the first questions that the lawyer will ask of a training organization that stopped the training of the litigant. Using the word “competency” and “competence” numerous times during their answer will not be an adequate defense. The lawyer will want to know the specific criteria that are objective, measurable,

transparent and fair, and which clearly demarcates the difference between competent and incompetent performance. As things currently stand, medicine would be in considerable difficulties. This is a very difficult issue to resolve.

Compounding this problem is our suspicion that the profession of medicine and the general public (and remember politicians and senior civil servants make up the general public as well) have contradictory notions about precisely what competence means. Medicine probably construes competency as something closer to the dictionary definition. In contrast, we believe that the general public's views of medical competence is something more akin to the dictionary definition of proficiency.

- *Competence*: describes those behaviors required for satisfactory (“threshold competence”) performance in a job
- *Proficiency*: describes the ability to perform a specific behavior (e.g., task) to the established performance standard in order to demonstrate mastery of the behavior; skillfulness in the command of fundamentals deriving from practice and familiarity

This is a relatively straightforward question to answer but the response may pose even more difficulties for medicine. It is our belief that the general public does not construe “medical competence” or just passing and no more. Medical competence appears to be construed as performing at a higher level. However, it would be useful if medicine could quantitatively answer this question and so avoid potentially awkward questions and possibly even more awkward answers. Damaging public confidence in medicine further is probably not a good idea at the present time!

Dreyfus et al. (1986) suggested that in the process of skill progression, there is never a clear demarcation between one level and the next (Chap. 8). This means, for example, that the performance characteristics of the novice will at certain times be more similar to the advanced beginner than they are to the novice level. This does not mean that at these times, the novice is a fully fledged advanced beginner. They may demonstrate some of the performance characteristics, but this is likely to be in superficial aspects such as technical skill and not in characteristics such as wisdom. This is most likely to be the case in surgical skill progression. For example, in the proficiency-based progression clinical trials that have already been conducted, the researchers would not argue that the proficiency-trained surgical trainees had the same procedural wisdom as the attending and consultant surgeons on whom their technical skill benchmark was based. All that the trainees did was demonstrate the proficiency benchmark of the more experienced surgeons on two consecutive training trials and having done this, they also demonstrated superior objectively assessed intraoperative performance than a traditional trained group. This means that the trainees demonstrated performance characteristics of proficient surgeons but this does not mean that the trainees themselves are proficient. Only one specific aspect of the performance was trained and tested during the clinical trial. We find the simplicity of this approach very appealing because it avoids convoluted discussions which have been ongoing for some period of time but at the same time does not compromise the quality of trainee performance.

**Fig. 12.1** A core competency satisfied, built on real-world defined attributes that are objectively assessable and based on empirically demonstrable characteristics which are defensible!

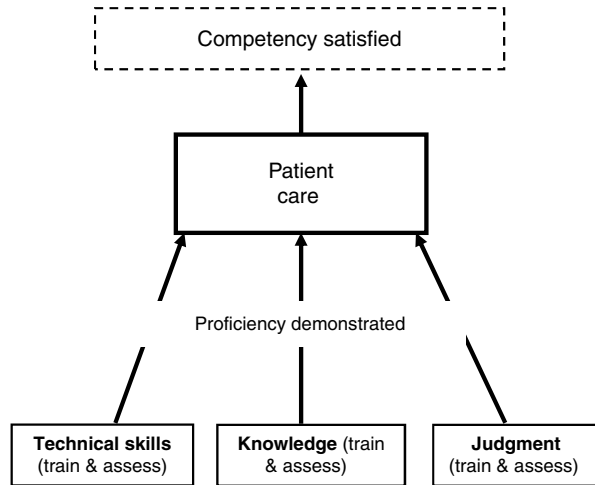
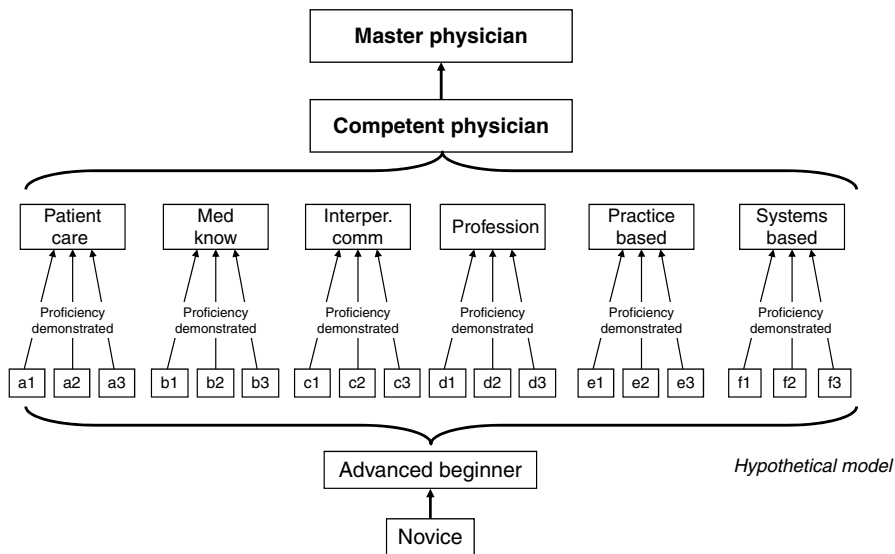


Figure 12.1 shows how this approach might be implemented in a manner similar to proficiency-based progression clinical trials that have already been conducted. It shows a hypothetical process of meeting the ACGME “Patient Care” core competency. Metric-based assessment for the constituent components for the patient care competency, i.e., technical skills, knowledge, and judgment could be developed very much in the manner we described in Chap. 5. These could then be used to characterize how experienced surgeons perform against these metrics, thus establishing a proficiency level. Trainees would then be required to demonstrate proficiency on the performance characteristics. Once demonstrated on all three performance characteristics, by default, the trainee has just demonstrated competency in this core competency. The precise number of times that the trainees should demonstrate proficiency or the methodology used for a trainee to demonstrate proficiency will still need some discussion but this is a relatively straightforward question that can be answered quantitatively. In its simplest form, the question asks: how many times must proficiency be demonstrated so that the trainee is assessed as safe as can be hoped for without significantly compromising the amount of time it takes to fully train a surgeon. Figure 12.2 shows how this approach might be applied to trainee surgeons demonstrating all six ACGME core competencies. After demonstrating proficiency in the different performance characteristics that constitute the core competency, the trainee is, by default, competent.

This approach to competency-based training avoids some of the difficulties of trying to operationally define competence in a way that the vast majority medical of practitioners will agree. It also ensures that there is no compromise in the quality of the skills the surgeon brings to patient treatment and care. It does however deal with the question of “What is the demarcation between competence and failure to reach competence”? Furthermore, it has answered the question of how to actually define what is competence. This approach is also flexible enough to allow the progression to be optimally paced for the trainee while still not compromising on the quality of



**Fig. 12.2** The ACGME six core competencies: An alternative hypothetical model for acquiring and demonstrating “medical competence”

training. Furthermore, it provides a very clear quantitative benchmark which has been unambiguously defined for trainees and potential litigants. The GMC in the UK has the assessment infrastructure already in place to implement such a strategy. They have formative assessments in the form of the DOPS and they have summative assessments in the form of PBAs (Chap. 8). The PBAs may need some development work so as to eliminate assessment items such as, “optimum” (without definition), “adequate,” “sound,” and “purposeful.” This is a relatively simple matter. They could then be used to quantitatively define levels of proficiency for the index procedures already identified. This would make a very robust assessment system.

Whatever approach is taken to solve the verification or falsifiability issue, a less ambiguous training endpoint will have to be developed by the major surgical training bodies around the world. As we have clearly indicated throughout this book, time in training is not a good predictor of skill and if, as the training bodies state, they have a competency-based training program, why not have competency or proficiency as the indicator of training completion rather than the time in training. If the trainees are given end of training benchmarks such as levels of proficiency that have been quantitatively defined on the basis of experienced surgeons performance, they will probably find that acceptable or very difficult to disagree with. Of course, this assumption is based on the premise that the training facilities are made available to them in order to demonstrate the level of proficiency. That means that they must have access to training facilities where they can engage in deliberate practice. Defining an unambiguous training endpoint could possibly create its own problems. For example, assuming that the issues pertaining to proficiency and competency are satisfactorily resolved with an unambiguous outcome, resulting in a clearly



defined quantitative end of training based on objectively assessed performance of the trainee, what are training bodies going to do with trainee surgeons who progress rapidly through the proficiency-based progression training cycle? It is assumed that proficiency will have been demonstrated in the process with something like their PBAs on real patients. Should the trainees who demonstrate proficiency first give up their operative cases so that their peers have more opportunities to demonstrate proficiency? or should they progress to the next training rotation? (Oh, but for such a problem)! This sort of scenario could play havoc with training rotations and the administration of a training program. However, it could also offer the opportunity to radically reduce the number of years in training without compromising the quality of the graduating surgeon.

## **Optimized Training Availability**

There is a growing body of data from clinical studies that shows proficiency-based progression training on simulation models is a better way to train procedural-based surgical skills. It is also clear that these training models work because they afford the opportunity for the trainee to engage in deliberate practice. Deliberate practice differs from repeated practice (the ethos of the traditional approach to training) because of the way the curriculum content is configured, delivered, and assessed. Trainees on a proficiency-based progression training schedule engage in deliberate practice with formative feedback, which shapes and optimizes their performance. The optimal application of this type of training program assumes that the trainee engages in a didactic educational program (which is also proficiency-based progression) before being offered technical skills training. The evidence from clinical trials shows that proficiency-based progression training using this approach has resulted in superior objectively assessed intraoperative performance when compared to the traditionally trained surgeons. These results have been demonstrated for basic laparoscopic procedures such as cholecystectomy and for advanced procedures such as laparoscopic Nissen fundoplication. The training for laparoscopic cholecystectomy was conducted on virtual reality simulation (Ahlberg et al. 2007; Seymour et al. 2002) and Nissen fundoplication training was completed on improvised simulation models (Van Sickle et al. 2008b). Deliberate practice coupled with formative feedback on both types of simulations significantly improved objectively assessed intraoperative performance in comparison to traditionally trained surgeons. However, one of the most important lessons learned from these studies was the additional effort that had to be invested to implement a proficiency-based progression training program on a simulation that was not computer generated. The simulation models used in the Van Sickle et al. study were perfectly adequate for achieving the goals of the training program and did a good job at facilitating the acquisition of intracorporeal suturing skills that transferred to intracorporeal suturing in Nissen fundoplication. The problem with this training program was the implementation of the formative and summative assessments. In vivo training on these simulation models

had to be supervised by a researcher who was very familiar with the application of the performance metrics. They also had to assess the quality on all of the knots tied during training. Possibly the process could have been made more efficient by training two subjects at a time rather than just one subject. However, even with this strategy, it is a very expensive approach to training; imagine a standard class size of 20–30 trainees. These are the sorts of numbers the Royal College of Surgeons in Ireland train daily in its skills laboratory.

One of the most important lessons learned from the Van Sickle et al. study was the value of computer-generated and scored virtual reality tasks. It makes the entire training process orders of magnitude more efficient. This is an important lesson because surgical training programs that opt to conduct this type of surgical training purely for training purposes and not for research purposes are very unlikely to have the personnel resources to invest. Although all of the researchers conducting the training in the Van Sickle et al. study were highly trained on the implementation of the formative metrics; the fact that they were delivered by a person rather than a computer can allow subjectivity to creep into the assessment process. Even if the researchers had implemented assessment of psychomotor performance using something like ICSAD for measuring hand movements, intra-operative task performance would still be required, during training, to comply with the formative assessment aspect of training. These findings and conclusions point to the urgent need for wider availability and use of computer-generated and scored virtual reality simulation tasks for training procedural-based skills such as surgery. Evidence clearly shows that they are effective and efficient at delivering deliberate practice training as part of a proficiency-based progression skill acquisition program. One of the problems that disciplines like surgery have is that most of the virtual reality simulations available commercially are for minimally invasive or endovascular procedures.

## **Open Surgical Simulation**

The traditional open incision remains the most common approach to performing surgical procedures. In spite of this, practically all of the surgical simulations that are currently available on the market are for some type of minimally invasive intervention such as laparoscopic, endoscopic, or endovascular. A range of silicone-based and animal tissue models are currently used for the training and assessment of surgical skills for open surgery. However, one of the problems with these tasks is that the silicone models vary in the degree they approximate the actual surgical task/procedure on a real patient. For example, some of the silicone models for training suturing are inappropriate for training a subcuticular suturing technique as the suture material tends to rip through the foam material. The bowel and anastomosis models also have similar problems, and while they may look acceptable when the task has been completed, they are really not very good for assessments such as leakage of the anastomosis. The water tends to seep through the small holes through which the

needle passed. The advantage of these types of models is that they can be used in almost any teaching space. Animal tissue can be used as an alternative to silicone with the advantage that in general, they have many of the properties of human tissue. However, the problem with these tasks is that they require specialist facilities for use and disposal, e.g., specialist tables, flooring, cleaning, etc. Both of these types of simulation training models have been used for training purposes for decades. However, with a better understanding of how to achieve effective and efficient training, e.g., deliberate practice, and pressures on the amount of time available for training, these models look increasingly unattractive. The greatest problem with using them is providing performance and summative feedback to the trainee in an efficient and cost-effective manner. Procedural-based medicine trainers and educationalists (undergraduate and postgraduate) should come to the realization that there is an urgent need to develop virtual reality simulations for the training of open surgical skills. There are some simulations that claim to train open surgical skills such as giving an intravenous injection or taking blood. There are also a number of fairly large projects which are underway around the world whose outputs look as though it would take relatively little effort to develop them into full-blown virtual reality simulations for open surgical skills. The Virtual Physiological Human project (<http://www.vph-noe.eu>) is an extension of the virtual human project that is trying to do a multi-scale model of the human body. There is also the 3D Anatomical Human (<http://3dah.miralab.ch>) which claims to be more aligned with real-time interactive simulations of humans for practical applications rather than the basic science focus of the VPH project. The Simulation Open Framework Architecture (SOFA) is an Open Source framework primarily targeted at real-time simulation, with an emphasis on medical simulation. It is mostly intended for the research community to help develop newer algorithms, but can also be used as an efficient prototyping tool (<http://www.sofa-framework.org/home>). These efforts are to be commended but the problem with these approaches is that they are mostly proof-of-concept systems or designed for research and development. Furthermore, showing high-quality anatomical images is all well and good for display purposes to show what is possible with virtual reality. The problems come when they are required to be used for interactive hands-on simulation training.

## Open Surgical Simulation: What Would It Take?

One of the major problems for the development of an open simulator is producing generalized solutions that are physics-appropriate and yet can run in real time. The more complex the interactions between tools, tissues, etc., the more complex the computation of the interactions become. For example, in a simple suturing task, the interactions will include needle holder grasping the needle that punctures tissue while needing to stabilize the tissue with another tool and also looping the suture thread around tools to begin cinching down the two sides of the wound which must contact one another and produce appropriate contact stresses and result in the proper

inversion/eversion of the wound sides. These highly complex interactions are currently being tackled by Dr. Dwight Meglan and Prof. Howard Champion's (Simquest, Silver Springs, USA) team in their construction of a simple open surgery wound closure task. One of the major problems in creating an open simulation that is physics based is that there are very few people in the world who are experienced at developing this technology for a real-world application.

To develop solutions for the problem of open surgical simulation will require a very focused effort and considerable developments in existing knowledge, including physics-based simulation and engineering. To create a generalize able model of an open surgical procedure interacting with the anatomy of a human, it should probably start with putting together measurement tools that can define exactly what is physically happening in surgery, e.g., movements of tools, forces/torques/pressures at the interface of tool-tissue, etc. Then catalogs need to be developed for all of the tissues that need to be simulated in terms of their mechanics and construction (heterogeneous materials like muscles, nerves, blood vessels, lymph ducts, etc.), and the like. Also included in this catalog would be how the tissues are interconnected. In addition, a catalog of all tool-tissue interactions, both in type as well as in mechanics – details like how grasp really happens (friction, mechanical interference, etc.), the process of tissue failure in cutting, ablation mechanics, etc. From these units of information, an engineering approach would need to be developed to construct various entities at a foundational level and form more complex tissues from these. As a lot of this information will be novel, quantitative engineering tests would need to be conducted at each level of construction to prove how well the simulations match reality, both in terms of mechanics as well as in terms of speed of computation. Simultaneous with the tissue buildup, detailed tool–tissue interaction physics would need to be developed and managed with the same approach for doing deconstructed simulations at the lowest level first and building up from those with the same engineering property and simulation assessments being conducted at each level.

To ensure optimal functionality, this project would require focus around one deliverable simulation project; large enough that it answers a lot of simulation, physics and, engineering questions about building an open surgical simulator, but is also something that was manageable. This would minimize the development of disparate entities with their own research and development agendas. To undertake this challenging project would require people who are good at computational numerics and who also appreciate the need for real-time results. It would also need people familiar with computed interactions because we have been informed that this turns out to be much harder than doing the physics of the objects (like finite elements) especially when you want to do it in real time. Obviously, an open surgical simulator would require haptics and graphics developers. Development would also require individuals who are comfortable at undertaking task deconstruction of the surgeries and defining an approach/architecture to build up a general solution. Finally, the simulation development would require high-level developers who would concentrate on the construction of the learning scenarios (tools and data), defining the learning focus of the scenarios, assembling some form of automated instruction/

mentoring as well as formative and summative metric aspects as well as verification and validation studies.

## Who Is Going to Pay?

A textbook on fundamental principles of surgical simulation would be incomplete if we did not attempt to address how the principles and practices that we have described and discussed are going to be implemented and paid for. If the decision is taken by a surgical training program to implement even part of a simulation-based deliberate practice regime for proficiency-based progression training, they are going to require more resources. The least that they will require is experienced assessors to ensure that trainees get sufficient formative feedback on their performance during training. This assumes of course that the program leaders have already conducted the task analysis, developed the intraoperative or task performance metrics, and validated them, including the development of proficiency levels. These developments will significantly improve the effectiveness of current training, particularly if they were coupled with an online didactic education program linked to the skills laboratory training and schedule in the appropriate order. The use of staff to provide performance assessment during training is not a particularly efficient approach. In the short term, we really do not see that medicine will have an alternative but to make the training of procedural skills more effective. Doing nothing is not a sensible option.

A more efficient approach would use computer-generated virtual reality tasks for training. Unfortunately, a virtual reality simulator for training open surgical skills does not currently exist. We have some idea of what it would take to develop an open surgical simulation platform (which we described above). The development of one simulation platform for a specific open surgical procedure would cost between £50 and £100 million and probably take 3–5 years to complete assuming that the appropriate expertise could be found and employed to build it. The development of virtual reality simulations which can be used as actual training devices is orders of magnitude more difficult than producing virtual reality images, no matter how sophisticated those images are. At the moment, it is not clear who will pay for the development of such a device. We shall come back to this issue after we had discussed funding for the simulators that currently do exist.

As we have pointed out on a number of occasions, virtual reality simulations currently exist for minimally invasive and endovascular procedure. Although these approaches still represent a minority of approaches to interventional procedural-based medicine, these types of procedures still constitute a substantial number of operations per year. Furthermore, these procedures are significantly more difficult to learn than the traditional open approach to surgery. Some of these simulators have been developed since the mid-1990s (e.g., MIST VR) and clinical data showing their effectiveness as training tools has been available since the start of the twenty-first century. There remains no consensus about who should pay for these

devices. In the USA, the ACGME has insisted that surgical training programs should provide access to simulations and simulators. Despite a relatively standardized training program in the USA, there appears to be no coherent approach to the purchase and implementation of surgical simulation. The American College of Surgeons launched a program to accredit institutions which aimed to enhance access to educational opportunities in surgical training (Haluck et al. 2007). No extra monies were available to fund accredited institutions even though it was acknowledged that the financial and logistical considerations of establishing an institution were considerable. One of the good things about this effort was that it was National with implicit agreement to share experiences (both good and bad) in relation to training and simulation. This approach at least ensures that institutions do not replicate the same mistakes.

Ironically, simulations for minimally invasive approaches to performing procedures are probably the easiest to fund. Medical device companies continue to refine and develop new instruments for performing surgical and other interventional procedures. Currently, most of the training that these companies conduct to ensure that the surgeon or physician are familiar with the instruments is conducted in animal laboratories. This is a very expensive way to train to use relatively straightforward devices. The medical device manufacturers who produce endovascular devices such as catheters, stents, and wires probably have the greatest incentive to use virtual reality simulations for training as the animal models that currently exist bare little similarity to operating on patients. Furthermore, training using full physics virtual reality simulations means that the doctor can be trained to use the exact same device, in the exact same order, on more or less the same anatomy as they would in a real patient. Although these companies have invested heavily in these devices, their attitude toward virtual reality simulations indicates that they are not really sure what a huge business opportunity full physics virtual reality simulation represents for them. This is probably because they do not fully understand the capabilities of full physics virtual reality simulation. Some of them may even believe that it does not look or feel like operating on a real patient. As we have explained in Chaps. 3 and 10, the sensations that individuals detect from operating on real patient human anatomy and surgical instruments are perceived differently by each individual and the function of virtual reality simulation is not to simulate each individual's perceptual experience, rather it is to provide a reference case that is anatomically correct which can facilitate completion of a full procedure using the exact same devices as on a real patient.

Virtual reality training is a less expensive way for device manufacturers to train their sales staff who in turn can provide training for doctors to use the device. We are surprised that more multinational medical device companies have not made greater use of full physics virtual reality simulation in the design and marketing of their product. Engine design, automobile manufacturers, and Formula 1 racing teams currently make extensive use of virtual reality simulation in the design and preparation of their products. We are not entirely sure of the budget ratio between marketing and manufacturing of a new medical device but we do know that it is substantial. It would seem to us that more aggressive use of virtual reality simulation

would give considerable manufacturing and marketing advantage which proportionately would almost certainly convert into increased sales. Furthermore, we would have thought that the FDA decision on including virtual reality training as part of the roll out of carotid artery stenting with embolic protection would have given a very clear lead on this issue (Gallagher and Cates 2004a).

Although medical device companies may not have used virtual reality simulation to its full potential in their research and development of a product, they certainly have been keen to sponsor training events that utilize simulation. At most of the major medical conferences for procedural-based disciplines such as surgery, interventional cardiology and interventional radiology, etc., virtual reality simulations are now a common sight in the booths of the large medical device manufacturers. There has been some discussion within the professional societies about approaching the large multinational medical device manufacturers and requesting that they pay (fairly large sums of money) for simulators for surgical training centers. However, even if the manufacturers paid for or “sponsored” the simulators, they would have no say on how the simulators would be used, nor of the curriculum content which from the manufacturers’ perspective may not seem a very good deal. As it currently stands, the medical device industry is relatively generous with its arm’s-length sponsorship of courses and events. However, the medical device industry continues to appear bemused by the potential of this very powerful technology. Paying for further original development of simulations does not appear to be imminent from this source.

The Royal College of Surgeons in Ireland has a well-developed surgical training center and pursues a training and assessment strategy using a wide range of simulations and simulators. They have also adopted one of the most innovative approaches to the implementation of a training and assessment strategy using simulation. Surgical trainees in Ireland *must* attend the national surgical training center for a minimum of 6 days per year for training. To pay for this facility, each trainee is charged €3,000 per year (which is tax deductible). However, this does not even cover 50% of what it costs to train trainee surgeons for the 6-days training provided. Irish surgical trainees would probably be considerably more disgruntled if they were charged in excess of €7,000 per year for their simulation-based training. The unwritten and unspoken understanding in postgraduate medical training in the USA seems to be that the trainee will work long hours, accept relatively poor pay and help to look after the attending surgeons’ patients in return for being trained as a surgeon. However, with reduced work hours and consequently reduced opportunities for training, especially in the operating room, this unwritten “arrangement” seems to be under increasing pressure. Furthermore, surgical trainees in the USA and mainland Europe simply can not afford the full costs of skills laboratory training.

One possibility that could be used to subsidize training within institutions is for attending/consultant surgeons to develop procedure-specific teaching modules that are accompanied with a fully developed didactic module and an edited video recording of a specific procedure with running commentary. For some operations such as endovascular procedures, the surgeon could also make available the patient-specific data that could be downloaded into a virtual reality simulator for the trainee to

practice the procedure that they had just studied. This for fee service could then be used by trainees (as well as much more experienced interventionists) to consolidate and expand their procedure experience. Indeed, whether or not this service develops commercially, we fully envisage it developing over the next decade to supplement the experience of experienced surgeons whose practice will probably be forced to become more and more specialized.

The possibility of professional societies and medical device manufacturers coming together to run and finance simulation-based training is currently a reality. Almost all of the procedural-based medical disciplines around the world rely heavily on the sponsorship of industry to help finance courses that they organize. This financial support reduces the cost of the courses but does not cover them completely. In general, this sponsorship is usually only available for trainees who are fairly advanced in their training or for consultant/attending courses. Furthermore, it seems highly likely that industry sponsorship for these types of courses in the future will become more and more restrictive as national governmental organizations and audit offices monitor ever closer the relationship between medical device companies and physicians. It is difficult to see what this relationship will morph into but we find it hard to believe that medical device manufacturers will not have a significant role in financing courses in the future. It may be that they sponsor or own the simulators on which the courses are run. The fact remains that interventional attending specialist courses must have hands-on experience with the devices they are going to use on real patients. Full physics virtual reality simulation certainly seems to us to be the best model on which to train and we do not see how training can be conducted without using the actual physical devices. Furthermore, it is probably best if an expert from the manufacturing company explains to the trainees how best to use the devices rather than have a surgeon or other interventionist explain how they use it. In our experience, these two accounts do not always correlate, and for safety and insurance purposes, it is probably best that the surgeon or interventionist hears directly from the manufacturers of the device how it should be used. Then, if the surgeon or interventionist decides not to use it the way suggested by the manufacturer, there can be little ambiguity where the fault lies if anything goes wrong.

An interesting development has been ongoing in Massachusetts at the Harvard Risk Management Foundation which provides malpractice insurance for doctors working in their health-care system. Anesthetists as well as obstetricians and gynecologists who have undergone a rigorous simulation and training program are eligible for up to 10% discount if they successfully complete the risk reduction course which involves team training simulation. Malpractice insurance for physicians in the USA is very expensive and a 10% reduction represents a substantial amount of money. We believe that the system could be optimized even further if the insurers insisted that course participants demonstrated a level of proficiency and that proficiency was based on the performance of a large group of interventionalists, e.g., surgeons, interventional cardiologists, and interventional radiologists. We believe that this would considerably reduce the risk of something untoward happening for the majority of physicians. We are very surprised that malpractice insurers have not made greater use of this facility particularly given the validation evidence



that currently exists and the very clear relationship between proficiency-based progression and improved intraoperative performance.

In the UK and Ireland, it is not uncommon for institutional changes of the magnitude that we have outlined here to be financed by central government. If we return to the issue of paying for the development of an open surgical simulator, none of the organizations that we have discussed thus far either have the resources or the inclination to invest in such a development. Instruments that are used to perform open surgery are (in general) not disposable and do not really change that much over the years; the simulation companies who could potentially develop an open simulator do not have £50–£100 million to invest; it is probable that the professional societies do not have that amount of spare cash lying around and even if they did, getting agreement from them as to which open surgical procedure should be simulated first would be an interesting exercise; anyway, surgical training has been conducted perfectly satisfactorily for centuries on real patients. None of these answers leads to a satisfactory state of affairs. The fact is that an open surgical simulator is urgently required. Even starting today, it would take 3–5 years to build a working prototype that could be copied. It would probably take another 5 years of concerted effort to get an open surgical simulator to the same level of fidelity that we have for endovascular interventions. Furthermore, there is a latent landmine waiting to explode. As interventional medicine becomes less and less invasive for more and more procedures, how are the surgical community expected to retain their expertise and skill level for open surgical procedures that are common today but will almost certainly become infrequent in the near future? Avoiding these difficult questions will not make them go away.

We believe that a number of fairly straightforward developments would clarify matters pertaining to the financing of training and simulation developments. Training systems in the USA, UK, and Ireland seem to agree that competency-based training programs are the way forward. However, the problem is that they cannot or would not agree on a quantitative definition of competency that is verifiable or falsifiable. Whether the training system is based on competency or proficiency may be considered a matter of semantics. An agreed-upon quantitative measure such as those that have been used in a number of studies and proposed here adds considerable clarity to the issue of how training should be conducted in the future. If a level of proficiency was mandated and training progression was dependent upon it, then organizations that run training courses would have a much clearer idea of the market they had to deal with. Proficiency-based progression training on a deliberate practice regime leads to superior intraoperative performance in comparison to traditional training; there can be little doubt about the data. It would be a very foolish pundit who would bet against these results translating into improved operative outcomes. This means that a number of national or regional training centers would be responsible for deliberate practice training regimes in skills laboratories. There would also be a National Curriculum with a coherent e-learning program which would also be proficiency-based and implemented as a pre-requisite for attending skills linked courses at the regional or national training center. Establishment of these centers would almost certainly have to be funded from governmental sources and where possible subsidized or co-sponsored by

industry. It should be remembered that the industry will want to use these facilities as well to train interventionists on their devices. This type of setup, with regional training centers possibly linked with an overarching informal organizational group such as the one set up by the American College of Surgeons would almost certainly ensure more efficient and effective training with national benchmarks.

In the financial year 2008/2009, Germany invested/spent €144 million, France €111 million, and the UK €107 million in The European Organization for Nuclear Research (Organisation Européenne pour la Recherche Nucléaire or originally *Conseil Européen pour la Recherche Nucléaire*) known as CERN. It is the largest particle physics laboratory in the world situated in the northwest suburbs of Geneva on the Franco–Swiss border (established in 1954). Each of these governments would argue that this money was invested/spent for the national and international good of mankind. In the USA, the Centers for Medicare and Medicaid Services (CMS) which is the equivalent of the NHS in the UK has an annual budget of approximately \$780 billion per year. In the UK, the Department of Health spent £100 billion in 2008/9. It is difficult to envisage why the finance necessary to fund the *proper* establishment of simulation and training centers cannot be found. The same is true about the development of an open surgical simulator.

As the development of new minimally invasive technologies are implemented into healthcare, it is very easy to forget that if something goes wrong, it is probably a surgeon performing an open surgical procedure who will have to pick up the pieces. The changes in work practices, the opportunities to acquire procedural expertise and wisdom are contracting dramatically. Furthermore, acquiring the basic surgical skills on real patients is no longer acceptable. Professionals in disciplines like surgery are now aware that the process of acquiring proficient skills can be made more effective and efficient with a regime of deliberate practice. However, the current curriculum needs to be reconfigured and new tools are required for the delivery of a newly configured curriculum. Simulation-based regional and national training centers that can deliver the curriculum are required urgently. These centers will not be cheap to establish and maintain. Furthermore, they need to be appropriately staffed as the absence of high-fidelity simulation that can provide formative feedback on performance must be substituted with experienced supervision and the application of the same metrics. The development of a full physics virtual reality simulator for training open surgical skills is extremely urgent and we would propose that it should be considered as a national or indeed international priority development in healthcare.

## Summary

Proficiency-based progression training on a simulator is a new approach to training doctors. Much of the ethos that is fundamental to proficiency-based progression training is not new. “Competency-based curriculum in any setting assumes that the many roles and functions involved in the doctor’s work can be defined and clearly expressed. It does not imply that the things defined are the

only elements of competence, but rather that those that can be defined represent a critical point of departure in curriculum development. Careful delineation of these components of medical practice is the first and most critical step in designing a competency-based curriculum” (McGaghie et al. 1978). Whether a training program is called competency-based progression or proficiency-based progression is a matter of semantics. However, as clearly stated by McGaghie et al., a training goal must be defined before it is established. No training program that currently claims to train competency-based progression has unambiguously defined competency endpoints that are falsifiable. In contrast, proficiency-based progression training studies have defined endpoints based on experienced surgeons’ performance and established clear endpoints that are verifiable and falsifiable.

Proficiency-based progression training works because of well-proven principles and practices of learning. To ensure the optimal effectiveness of a proficiency-based progression training program does not require a radical change in the current curriculum content. However, what does require radical change is how that curriculum is delivered and implemented. Virtual reality simulation is a very powerful training tool for the delivery of deliberate practice coupled to formative and summative metrics on performance. In the absence of computer-generated simulation, formative metrics on training performance needs to be delivered by a trainer who is very experienced at performance assessment. Some virtual reality simulators currently exist in minimally invasive surgery and endovascular procedures. There are none for the training of open surgical procedures despite the fact that open surgery remains the most common type of procedural intervention and is also associated with the highest rate of errors. This situation needs to be addressed urgently.

A training program that has a clear end point must provide the facilities and opportunities for learning to meet the level of proficiency. A deliberate practice training regime affords the opportunity for independent pacing of skill acquisition; a coherent curriculum with appropriately sequenced learning material; and a variety of learning experiences (lecturers, seminars, small group teaching, e-learning, silicon models, virtual reality emulators, high-fidelity virtual reality simulators and real patients) optimize learning availability; and formative and summative metric-based assessments maximize the probability of learning. Although this approach to medical education and training may be conceptually and intellectually appealing, it represents a paradigm shift in how doctors are educated and trained.

## References

- Ahlberg G, Enochsson L, Gallagher AG, et al. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *Am J Surg*. 2007;193(6):797-804.

- Bann S, Kwok KF, Lo CY, Darzi A, Wong J. Objective assessment of technical skills of surgical trainees in Hong Kong. *Br J Surg*. 2003;90(10):1294-1299.
- Beall DP. The ACGME institutional requirements: what residents need to know. *J Am Med Assoc*. 1999;281(24):2352.
- Bower JL, Christensen CM. Disruptive technologies: catching the wave. *Harv Bus Rev*. 1995;73:43-53.
- Centres R. Cholecystectomy practice transformed. *Lancet*. 1991;338(8770):789-790.
- Dawson S. Procedural simulation: a primer I. *Radiology*. 2006;241(1):17.
- Dreyfus HL, Dreyfus SE, Athanasiou T. *Mind Over Machine*. New York: Free Press; 1986.
- Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. *Psychol Rev*. 1993;100(3):363-406.
- Gallagher AG, Cates CU. Approval of virtual reality training for carotid stenting: what this means for procedural-based medicine. *J Am Med Assoc*. 2004a;292(24):3024-3026.
- Gallagher AG, Cates CU. Virtual reality training for the operating room and cardiac catheterisation laboratory. *Lancet*. 2004b;364(9444):1538-1540.
- Gallagher AG, Cowie R, Crothers I, Jordan-Black JA, Satava RM. PicSOR: an objective test of perceptual skill that predicts laparoscopic technical skill in three initial studies of laparoscopic performance. *Surg Endosc*. 2003a;17(9):1468-1471.
- Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. *Surg Endosc*. 2003b;17(10):1525-1529.
- Gallagher AG, Smith CD, Bowers SP, et al. Psychomotor skills assessment in practicing surgeons experienced in performing advanced laparoscopic procedures. *J Am Coll Surg*. 2003c;197(3):479-488.
- General Medical Council. *Tomorrow's Doctors: Recommendations on Undergraduate Medical Education*. London: GMC; 1993.
- Haluck RS, Satava RM, Fried G, et al. Establishing a simulation center for surgical skills: what to do and how to do it. *Surg Endosc*. 2007;21(7):1223-1232.
- Kazdin AE. *Behavior Modification in Applied Settings*. Pacific Grove: Brooks/Cole Publishing Co.; 1994.
- Khan M, Bann S, Darzi A, Butler P. Assessing surgical skill using bench station models. *Plast Reconstr Surg*. 2007;120(3):793-800.
- Kohn LT, Corrigan JM, Donaldson MS. *To Err Is Human: Building a Safer Health System*. Washington DC: National Academy Press; 2000:196-197.
- Larsen CR, Grantcharov T, Aggarwal R, et al. Objective assessment of gynecologic laparoscopic skills using the LapSimGyn virtual reality simulator. *Surg Endosc*. 2006;20(9):1460-1466.
- Martin P, Martin PR, Bateson P. *Measuring Behaviour: An Introductory Guide*. Cambridge: Cambridge University Press; 1993.
- Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. 1997;84(2):273-278.
- Mayo WJ. Medical education for the general practitioner. *Journal of the American Medical Association*, 1927;88(18):1377-1378.
- McGaghie WC, Sajid AW, Miller GE, Telder TV, Lipson L. *Competency-based curriculum development in medical education: an introduction; Competency-based curriculum development in medical education: an introduction*. Geneva: World Health Organization 1978.
- Popper KR. *Objective Knowledge: An Evolutionary Approach*. Oxford: Clarendon Press; 1979.
- Reason J. Human error: models and management. *Br Med J*. 2000;320(7237):768.
- Regenbogen SE, Greenberg CC, Studdert DM, Lipsitz SR, Zinner MJ, Gawande AA. Patterns of technical error among surgical malpractice claims: an analysis of strategies to prevent injury to surgical patients. *Ann Surg*. 2007;246(5):705.
- Senate of Surgery. *Response to the General Medical Council Determination on the Bristol Case: Senate Paper 5*. London: The Senate of Surgery of Great Britain and Ireland; 1998.

- Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg.* 2002;236(4):458-463; discussion 463-454.
- Somville F, van Sprundel M, Somville J. Analysis of surgical errors in malpractice claims in Belgium. *Acta Chir Belg.* 2010;110(1):11-18.
- Swets JA. *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics: Collected Papers.* Mahwah: Lawrence Erlbaum Associates; 1996.
- Van Sickle K, Baghai M, Huang IP, Goldenberg A, Smith CD, Ritter EM. Construct validity of an objective assessment method for laparoscopic intracorporeal suturing and knot tying. *Am J Surg.* 2008a;196(1):74-80.
- Van Sickle K, Ritter EM, Baghai M, et al. Prospective, randomized, double-blind trial of curriculum-based training for intracorporeal suturing and knot tying. *J Am Coll Surg.* 2008b;207(4):560-568.