

● 2003 INVITED ADDRESS

Deliberate Practice and the Acquisition and Maintenance of Expert Performance in Medicine and Related Domains

K. ANDERS ERICSSON

The factors that cause large individual differences in professional achievement are only partially understood. Nobody becomes an outstanding professional without experience, but extensive experience does not invariably lead people to become experts. When individuals are first introduced to a professional domain after completing their education, they are often overwhelmed and rely on help from others to accomplish their responsibilities. After months or years of experience, they attain an acceptable level of proficiency and are able to work independently. Although everyone in a given domain tends to improve with experience initially, some develop faster than others and continue to improve during ensuing years. These individuals are eventually recognized as experts and masters. In contrast, most professionals reach a stable, average level of performance within a relatively short time frame and maintain this mediocre status for the rest of their careers. The nature of the individual differences that cause the large variability in attained performance is still debated. The most common explanation is that achievement in a given domain is limited by innate factors that cannot be changed through experience and training; hence, limits of attainable performance are determined by one's basic endowments, such as abilities, mental capacities, and innate talents. Educators with this widely held view of professional development have focused on identifying and selecting students who possess the necessary innate talents that would allow them to reach expert levels with adequate experience. Therefore, the best schools and professional organizations nearly always rely on extensive testing and interviews to find the most talented applicants. This general view also explains age-related declines in professional achievement in terms of the inevitable reductions in general abilities and capacities believed to result from aging.

In this article, I propose an alternative framework to account for individual differences in attained professional development, as well as many aspects of age-related decline. This framework is based on the assumption that acquisition of expert performance requires engagement in deliberate practice and that continued deliberate practice is necessary for maintenance of many types of professional performance. In order to contrast this alternative framework with the traditional view, I first describe the account based on innate talent. I then provide a brief review of the evidence on deliberate practice in the acquisition of expert performance in several performance domains, including music, chess, and sports. Finally, I review evidence from the acquisition and maintenance of expert performance in medicine and examine the role of deliberate practice in this domain.

The Traditional View of Skill Acquisition and Professional Development: History and Some Recent Criticisms

The traditional view of skill acquisition is based on the assumption that innate biological capacities limit the level of achievement that a person can attain. Sir Francis Galton is generally recognized for developing the scientific basis for this view in the 19th century. In his pioneering book, *Hereditary Genius*,¹ he presented evidence that height and body size was determined genetically, and most impor-

tantly, he argued that similar innate mechanisms must determine mental capacities, stating:

Now, if this be the case with stature, then it will be true too as regards every other physical feature—as circumference of head, size of brain, weight of grey matter, number of brain fibres, &c.; and thence, a step on which no physiologist will hesitate, as regards *mental capacity* [italics added].¹, pp. 31–2

At the same time, Galton clearly acknowledged the effects of practice and the need for training to reach high levels of performance in any domain. However, he argued that improvements are rapid only in the beginning of training and that subsequent increases become increasingly smaller, until “Maximal performance becomes a rigidly determinate quantity.”¹, p. 15 According to Galton, the relevant heritable capacities set the upper bound for an individual's physical and mental achievements, and once all of the training benefits have been attained through sufficient practice, then the immutable limit for performance is reached that “Nature has rendered him capable of performing.”¹, p. 16 According to Galton, the immutable characteristics that limit maximal performance cannot be altered through training. By extension, they must have been innately endowed. Galton's arguments for the importance of innate factors in elite performance were quite compelling and, thus, have had a lasting impact on our culture's view of ability and expertise.

Contemporary theories of skill acquisition^{2,3} are consistent with Galton's general assumptions and with the observations on the course of professional development. When individuals are first introduced to an activity such as driving a car or playing golf, their primary goal is to reach a level of mastery that will allow them to perform everyday tasks at an acceptable level or to engage proficiently in recreational activities with their friends. During the first phase of learning,² novices try to understand the activity and concentrate on avoiding mistakes. With more experience in the middle phase of learning, gross mistakes become increasingly rare, performance appears smoother, and learners no longer need to concentrate as hard to perform at an acceptable level. After a limited period of training and experience—frequently less than 50 hours for most recreational activities such as typing, playing tennis, and driving a car—an acceptable standard of performance is typically attained. As individuals adapt to a domain and their performance skills become automated, they are able to execute these skills smoothly and without apparent effort. As a consequence of automation, performers lose conscious control over execution of those skills, making intentional modifications difficult. Once the automated phase of learning has been attained, performance reaches a stable plateau with no further improvements, which is consistent with Galton's assumption of a performance limit.

The principal difference between acquisition of everyday skills and professional development appears to be primarily a difference in time scale. Whereas proficiency in everyday skills is attained rapidly, professional development (including the prerequisite educa-

tion) is completed only after years or even decades of experience. In their seminal work, Simon and Chase⁴ argued that in order to attain an international level of performance in chess, an individual must maintain full-time involvement in the activity for at least ten years. Their research on chess masters' memories for regular game positions suggested that the masters had acquired some 50,000 chunks or patterns, and they highlighted the parallels between reaching this highly skilled performance level and acquiring a language with its large vocabulary. Simon and Chase⁴ proposed a theory of expertise where future experts gradually acquired patterns and knowledge about how to react in situations as a direct consequence of their continued experience in the domain. Based on the observation that most people are able to master their first language after many years of experience even without formal instruction, scientists started considering the possibility that sufficient experience (over ten years of full-time engagement) might automatically lead to expertise in a performance domain. Eventually, some scientists viewed sufficient length of experience in a domain (over ten years) as a reliable indicator of expertise.

Several reviews over the past decade⁵⁻⁷ have shown that empirical evidence for the traditional views regarding the development of expertise through extended experience alone is surprisingly limited. First and most surprisingly, the performance of experts, who are nominated by their peers based on their extensive experience and reputation, is occasionally unexceptional for representative tasks from their domain of expertise. For example, highly experienced computer programmers' performance on programming tasks is not always superior to that of computer science students,⁸ and physics professors from UC Berkeley were not always consistently superior to students on introductory physics problems.⁹ More generally, the level of training and experience has frequently been only weakly linked to objective measures of performance in a domain. For example, the length of training and professional experience of clinical psychologists is not related to their efficiency and success in treating patients,¹⁰ and extensive experience with software design is not associated with consistently superior proficiency on presented tasks.^{11,12} Similarly, when "wine experts" are required to detect, describe, and discriminate characteristics of a wine without knowledge of its identity (seeing the label on the bottle), their performance is only slightly better than those generated by regular wine drinkers.^{13,14} More generally, reviews of decision making^{15,16} show that experts' decisions and financial advice on investing in stocks have surprisingly low accuracy that does not improve with additional experience. Similar phenomena have been documented in several other areas of expertise.⁶ In the second half of this article, I will review evidence on whether highly experienced doctors, with reputations as experts, display consistently superior performance to their less experienced colleagues and, if so, I will identify the characteristics of those medical activities that reveal such superiority.

Second, the traditional views of professional development and skill acquisition assume that people will reach a stable asymptotic level of performance after sufficient experience. This assumption is inconsistent with the documented ability of highly experienced individuals to continue to improve their performance through training. Research has shown that when even highly experienced workers and professionals are appropriately motivated, they are able to improve their objective performance, sometimes dramatically.⁵ Finally, many efforts have been made to measure children's, adolescents', and young adults' basic capacities, such as short-term memory and attention, as well as visual, auditory, spatial, and motor skills, to assess the possible innate limits of their attainable performance. Efforts to use these estimated capacities to predict adult professional achievement have been disappointing and largely unsuccessful.⁶ Efforts to identify innate individual characteristics that are critical to attaining expert performance yet are resistant to modification by extensive training have not proved fruitful.^{17,18} The only such innate characteristics for which the genetic difference between people and the associated mechanisms are well

understood are body size and height,^{19,20} wherein above-average height provides an advantage in basketball, while below-average height facilitates elite performance in gymnastics.

These findings raise doubts about the common-sense view of genetically determined limits that severely constrain people's attainable performance. They suggest that highly motivated individuals should be able to influence their attained performance levels to a much greater degree than is traditionally assumed. The remainder of this article addresses two broad objectives. First, emerging findings are presented on how performers attain expert levels of achievement in diverse domains, including music, sports, and chess. A particular focus here is on the role of deliberate practice and training quality in mediating improvements in current levels of performance. Second, the expert-performance approach, with its emphasis on deliberate practice, is applied to research on expertise in medicine. Here, insights gained from the study of expertise in other areas provide interesting ideas for enhancing and maintaining attained levels of performance in the medical domain.

The Scientific Study of Expert Performance and Its Acquisition

In the introduction to this article, several examples were given where respected experts, such as wine connoisseurs and famous stockbrokers, failed to demonstrate consistently superior performances on representative tasks, such as investing money in stocks and describing wines. Scientists can no longer assume that individuals reputed to be experts based on their extended experience will display the superior achievement indicative of expertise in a domain. Ericsson and Smith²¹ criticized studies of expertise that merely looked for differences between "experts" (defined by social criteria) and less experienced individuals. They proposed instead that researchers focus on superior performance in a domain and identify any individual who consistently exhibits superior performance—whether they are socially recognized as expert or not. The first step is to construct representative tasks that capture the essence of expertise in the domain where the superior performer can exhibit their superior performance in a consistent and reproducible manner.

The focus on reproducible scientific evidence on superior performance provides a framework for evaluating anecdotes about the achievements of athletes, musicians, and scientists. When the "hard" scientific evidence for the most amazing achievements is scrutinized, most of these incidents cannot be substantiated by independent and unbiased sources.²⁰ Often, the only sources of these anecdotes are the exceptional persons telling stories about their own childhood achievements when interviewed as adults. In other cases, the individuals observing the event may have misinterpreted what actually happened. For example, when a golfer sinks a 40-foot putt to win a golf tournament, it is often assumed that they did so because of their amazing ability.²² However, scientists have analyzed elite golfers' putting accuracy at different distances during tournament conditions and conducted experiments wherein golfers are asked to make the same shot ten to 20 times in row. The results of these analyses show that the consistency of their shots is never perfect, though much higher than that of less skilled golfers.²² Even the paths of the experts' properly played shots are influenced by random factors beyond the golfers' control and, thus, sinking long putts is due in large part to chance factors, even for the best golfers. To build a science of exceptional performance, we need to restrict the scientific evidence to those aspects of phenomena that can be repeatedly and reliably observed, such as the expert golfers' reduced variability in outcomes of shots. Ideally, from an empirical viewpoint, one would reproduce the everyday phenomenon of superior achievement in the laboratory, so it can be examined under standardized and experimental conditions.

In many domains, it is possible to measure expert performance by observing elite performers as they reproduce their superior achievement under controlled laboratory conditions.²¹ Many types of superior performance by experts are reproduced repeatedly in ev-

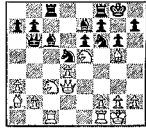


Domain	Presented Information	Task
Chess		Select the best chess move for this position
Typing		Type as much of the presented text as possible within one minute
Music		Play the same piece of music twice in same manner

Figure 1. Three examples of laboratory tasks that capture the consistently superior performance of domain experts in chess, typing, and music. (From "Expertise," by K. A. Ericsson and Andreas C. Lehmann, 1999, *Encyclopedia of Creativity*. Copyright by Academic Press.)

everyday life. For example, elite runners who finish the mile in less than four minutes can reproduce their exceptional running times repeatedly at different competitions. In sports there is a long tradition of creating fair competition by designing standardized comparable conditions for all participating athletes. The same is true for competitions in music, dance, and chess. In all of these domains, elite individuals consistently outperform their less accomplished counterparts.

Expert performers have trained to be able to reproduce their superior performance under representative conditions in everyday life whenever it is required during competition and training. Ericsson and Smith²¹ described how it is possible to identify naturally occurring activities that capture the essence of expertise in a given domain. For example, in his pioneering work on chess expertise, de Groot²³ argued that the essential task for expert chess playing consists of selecting the best moves for critical positions during chess games. To study performance, de Groot extracted chess positions from games between chess masters and set up a controlled laboratory situation where he could present these positions one at a time to an individual chess player (see Figure 1). The chess players were instructed to think aloud while they selected the best move for the presented position. Ability to select the best chess moves under these conditions is closely correlated with ratings for tournament competitors.²⁴

Another example is found in the study of expert performance in typing. Given that expertise in typing should generalize to any kind of textual content, we can simply give all typists the same text material and ask them to type it accurately as fast as possible. The final example given in Figure 1 illustrates a common obstacle to the study of expertise, namely that experts can excel at a task that less skilled individuals are unable to complete. In the study of music expertise, for example, we are confronted with the problem that the expert musicians typically perform pieces of music that are too difficult for less accomplished musicians to master. It is, however, possible to instruct all musicians to play familiar or unfamiliar pieces of lower difficulty and then ask them to repeat their own performances in as much detail as possible. When asked to do so, experts can repeat their performances with much less deviation from their original renditions than can less skilled musicians, thus exhibiting greater control over their performance. In related approaches, snooker players can be asked to make several shots for each of a series of fixed configurations of billiard balls, and golfers can be asked to make several putts for each of many ball locations on a green.²² In sum, it is possible to identify a collection of representative tasks that capture the essence of expertise in most domains,

and those tasks can be administered to all participants under controlled and standardized conditions. In the latter portion of this article, I will identify tasks that simulate the conditions for the essential activities in a few different domains in medicine, such as diagnosis and surgical treatments.

The Acquisition of Superior Reproducible (Expert) Performance

Having confined our attention to superior, reproducible performance that can be measured, it is possible to identify several claims about the acquisition of expertise that generalize across different domains.^{6,20} First, longitudinal assessments of performance reveal that performance increases gradually and that there is no evidence for abrupt increases from one trial to the next. Even when the performance of child prodigies in music and chess are measured by adult standards, the increases in achievement are found to be gradual and extended over many years. Second, the age at which experts typically reach their peak career performance is the middle to late 20s for many vigorous sports, and one or two decades later (i.e., in the 30s and 40s) for less physically intense games, as well as the arts and sciences. This continued, often extended, development beyond age 18 implies that the best individuals are able to engage in practice activities that lead to further improvements in performance even after physical maturation is reached. Finally, all performers, including the most "gifted" or "talented," need a minimum of approximately ten years of intense involvement before they reach an international level in sports, sciences, and the arts.^{4,5} Most elite performers, in fact, take considerably longer to reach that level. This ten-year rule of required engagement in domain-related activities is the most compelling evidence for the necessity of experience to attain high levels of performance.

From interviews with international-level performers in several domains, Bloom and his colleagues found that elite performers are nearly always introduced to their future domains of expertise as young children in a playful manner. As soon as they reveal enjoyment for the activity and show promise compared to their local peers, they receive help both in seeking out a teacher or coach and to begin regular training. Based on these interviews, Bloom²⁵ argued that access to superior training resources during development is necessary to reach the highest levels.

The best evidence for the value of current training methods and intensive schedules of practice comes from historical comparisons.^{22,26} The most dramatic improvements in the level of performance over historical periods are found in sports and are associated with improved quality and quantity of practice. Contemporary elite athletes' performance is much superior to the gold medal winners of the early Olympic Games.⁵ In some events where performance can be measured objectively (i.e., with times or distances), current winning performances are as much as 30–50% better.²⁷ Such a drastic improvement indicates the superiority of contemporary training methods, beyond what could be obtained through technological advances and better equipment, such as lighter running shoes, alone.

To further explore the role of practice in attaining expert levels of performance, Ralf Krampe, Clemens Tesch-Römer, and I⁵ tried to identify those training activities that were most closely associated with consistent improvements in performance and referred to them as deliberate practice. From a review of studies of learning and skill acquisition, we found evidence for consistent gradual improvement of performance when the following conditions were met. First, the participants were instructed to improve some aspect of performance for a well-defined task. Second, they were able to get detailed immediate feedback on their performance. Finally, they had ample opportunities to improve their performance gradually by performing the same or similar tasks repeatedly. The participants were able to keep improving their performance during extended training as long as the training sessions were limited to around an hour—a time that typical college students are able to maintain sufficient concentra-

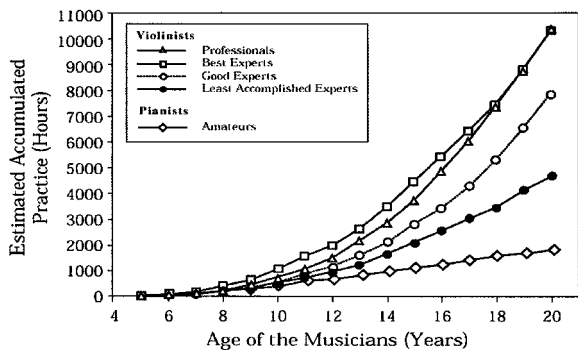


Figure 2. Estimated amount of time for solitary practice as a function of age for the middle-aged professional violinists (triangles), the best expert violinists (squares), the good expert violinists (empty circles), the least accomplished expert violinists (filled circles) and amateur pianists (diamonds). (From “The role of deliberate practice in the acquisition of expert performance,” by K. A. Ericsson, R. Th. Krampe, and C. Tesch-Römer, 1993, *Psychological Review*, 100(3), p. 379 and p. 384. Copyright 1993 by American Psychological Association. Adapted with permission.)

tion to sustain active efforts to improve. These deliberate efforts to increase one’s performance beyond its current level involve problem solving and finding better methods to perform the tasks. Engaging in practice activities with the primary goal of improving some aspect of performance is an integral part of deliberate practice.

The importance of deliberate practice in attaining expert performance was first demonstrated in a study of expert musicians studying at a famous music academy in Berlin. Three groups of the expert musicians who differed from each other in level of attained music performance were selected. All of the expert musicians were interviewed about how they spent their daily lives and were asked to keep detailed diaries of their activities for a week. Although all expert musicians were found to spend a similar amount of time when all types of music-related activities were combined, the two best groups of expert musicians were found to spend more time in solitary practice. When the experts practiced by themselves, they concentrated on improving specific aspects of the music performance as directed by their music teachers, thus meeting the criteria for deliberate practice. The best groups of expert musicians spent around four hours every day, including weekends, in this type of solitary practice. From retrospective estimates of practice, Ericsson et al.⁵ calculated the number of hours of deliberate practice that the three groups of musicians, along with two reference groups, had accumulated by a given age (see Figure 2). By the age of 20, the group of the best expert musicians (along with a reference group of musicians belonging to international orchestras) had spent over 10,000 hours of practice, which is 2,500 and 5,000 hours more than the two less accomplished groups of expert musicians and 8,000 hours more than a reference group of amateur pianists of the same age.²⁸

Several studies and reviews have found a consistent association between the amount and quality of solitary activities meeting the criteria of deliberate practice and performance in chess,²⁹ in music,^{28,30,31} and in different types of sports.^{32–34} The concept of deliberate practice also accounts for many earlier findings in other domains,⁶ as well as for the results of the rare longitudinal study of elite athletic performers.³⁵

Deliberate practice has even been found to be a key factor in maintaining expert levels as performers reach older ages. Although the performance of most professionals decreases, there are a few intriguing exceptions. A sufficient amount of weekly deliberate practice has been shown to allow expert pianists in their 50s and 60s to maintain their piano performances at a comparable level to that of young experts, although the older musicians displayed normal age-related declines on standardized tests.²⁸ Similarly, older masters

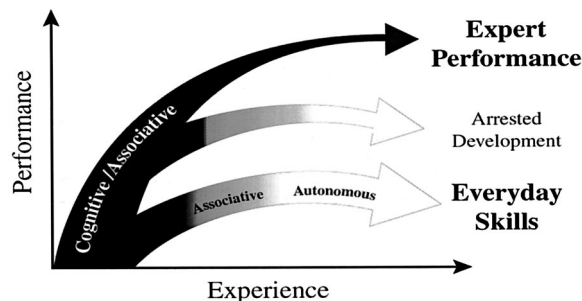


Figure 3. An illustration of the qualitative difference between the course of improvement of expert performance and of everyday activities. The goal for everyday activities is to reach as rapidly as possible a satisfactory level that is stable and “autonomous.” After individuals pass through the “cognitive” and “associative” phases, they can generate their performance virtually automatically with a minimal amount of effort (see the gray/white plateau at the bottom of the graph). In contrast, expert performers counteract automaticity by developing increasingly complex mental representations to attain higher levels of control of their performance and will therefore remain within the “cognitive” and “associative” phases. Some experts will, at some point in their career, give up their commitment to seeking excellence and thus terminate regular engagement in deliberate practice to further improve performance, which results in premature automation of their performance. (Adapted from “The scientific study of expert levels of performance: general implications for optimal learning and creativity” by K. A. Ericsson in *High Ability Studies*, 9, p. 90. Copyright 1998 by European Council for High Ability.)

in the game GO are able to maintain their performance and related skills,³⁶ and master athletes show the key importance of continued intense physical training.^{19,37} The age-related decreases in performance appear to result primarily from reductions of regular deliberate practice, rather than as a direct consequence of aging per se.³⁸

Complex Mechanisms That Mediate Expert Performance and Continued Learning

The fundamental theoretical challenge is to explain how most people and professionals reach a stable performance asymptote within a limited time period, whereas the expert performers are able to keep improving their performance for years and decades. When people and professionals are first introduced to an activity, their primary goal is to reach a sufficient level of mastery that is acceptable to other people in the domain. According to the traditional theory of skill acquisition,² people need initially to concentrate on what they are going to do in order to reduce gross mistakes, as illustrated in the lower arm of Figure 3. With more experience, their salient mistakes become increasingly rare, their performance appears smoother, and they no longer need to concentrate as hard to perform at an acceptable level. After some limited training and experience—frequently less than 50 hours for most recreational activities such as skiing, tennis, and driving a car—an acceptable standard of performance is attained without much need for effortful attention. As individuals’ behaviors are adapted to the performance demands and become increasingly automated, they lose conscious control and are no longer able to make specific intentional adjustments.

In direct contrast, expert performance continues to improve as a function of more experience, coupled with deliberate practice. The key challenge for aspiring expert performers is to avoid the arrested development associated with automaticity and to acquire cognitive skills to support their continued learning and improvement. The expert performer counteracts the tendencies toward automaticity by actively acquiring and refining cognitive mechanisms to support continued learning and improvement, as shown in the upper arm of Figure 3. The experts deliberately construct and seek out training situations in which the desired goal exceeds their current level of performance. They acquire mechanisms that are designed to increase their control and ability to monitor performance in repre-

sentative situations from the domain of expertise.^{7,22,39} For example, a chess player acquires improved memory skills to support working memory during the planning of the consequences of alternative moves for a chess position. In fact, chess masters are able to play chess “blindfold” without even seeing the chess board.⁴⁰ Expert typists look further ahead in the text and are thus able to prepare future keystrokes ahead of time to increase their typing speed.⁴¹ The rapid reactions of expert athletes are not due to greater speed of their nerve signals, but depend rather on their ability to better anticipate future situations and events by reactions to advanced cues.⁴² For instance, expert tennis players are able to anticipate where a tennis player’s serves and shots will land even before the player’s racquet has contacted the ball.⁴³

What kind of deliberate practice could possibly lead experts to keep improving their cognitive representations and mechanisms that mediate their superior performance? It is not obvious how an advanced chess player, who can easily beat all other players in the chess club, can improve in this unchallenging environment. How is it possible to improve one’s ability to plan and to select the best action in a given game situation? Chess players typically solve this problem by studying published games between the very best chess players in the world. They play through the games one move at a time to determine if their selected move will match the corresponding move originally selected by the masters. If the chess master’s move differed from their own selection, this would imply that their planning and evaluation must have overlooked some aspect of the position. By more careful and extended analysis, the chess expert is generally able to discover the reasons for the chess master’s move. Serious chess players spend as much as four hours every day engaged in this type of solitary study.^{5,29} By spending a longer time analyzing the consequences of moves for a chess position, players can increase the quality of their decisions. With more study, individuals refine their representations and can access or generate the same information faster.

Extensive research on typing provides some of the best insights into how speed of performance can be increased through deliberate practice that refines the representations mediating anticipation. The key finding is that individuals can systematically increase their typing speed by exerting themselves as long as they can maintain full concentration, which is typically only 15–30 minutes per day for untrained typists. While straining themselves to type at a faster rate—typically around 10–20% faster than their normal speed—typists seem to strive to anticipate better, possibly by extending their gaze further ahead.

The faster tempo also serves to uncover keystroke combinations in which the experts are comparatively slow and less efficient. By successively eliminating weaknesses, typists can increase their average speed and practice at a rate that is still 10–20% faster than the new average typing speed. The general approach of finding methods to push performance beyond its normal level—even if that performance can be maintained only for short time—offers the potential for identifying and correcting weaker components and enhancing anticipation that will improve performance.

Once we conceive of expert performance as mediated by complex integrated systems of representations for the execution, monitoring, planning, and analyses of performance, it becomes clear that its acquisition requires an orderly and deliberate approach. Deliberate practice is therefore designed to improve specific aspects of performance in a manner that assures that attained changes can be successfully integrated into representative performance. Hence, practice aimed at improving integrated performance cannot be performed mindlessly or independent of the representative context for the target performance. More accomplished individuals in the domain, ideally professional coaches and teachers, will play an essential role in guiding the future experts to acquire superior performance in a safe and effective manner.

In sum, the expert-performance approach has demonstrated how to capture superior expert performance in the laboratory, and

experimental studies have uncovered the complex mechanisms that mediate this level of superior achievement. These mechanisms are not the result of mere experience. Rather, their development requires engagement in deliberate practice that gradually builds complex integrated systems of representations for the execution, monitoring, planning, and evaluation of actions.^{6,7,22,39,44,45} When expert performers stop engaging in deliberate practice, their current performances tend to become automated, and development of structures is prematurely arrested (see the middle arm of Figure 3). In the next section, I will attempt to apply this framework to examine the acquisition, training, and maintenance of superior performance in medicine.

Expert Performance in Medicine

In most professional domains, including medicine, it takes a relatively long time for students to acquire the relevant knowledge and skills required for the profession. There is also a long period of supervised training where less experienced professionals gradually take on increased responsibility for the essential tasks in the domain, such as treating patients. Even after a physician is licensed to practice medicine, there is often continued training and experience required to become a specialist and eventually gain recognition as an expert. In medicine, there is general agreement regarding one’s level of attained expertise, as well as which individuals are considered experts. Most studies on medical expertise have recruited medical doctors and students based on their socially recognized levels of expertise and length of experience. It is common to identify five stages of learning as a function of instruction and experience, where a novice has to go through three intermediate stages before they are capable of reaching the ultimate stage as an intuitive expert.⁴⁶ It is generally assumed that these stages would correspond closely to individuals’ observed performance on representative tasks that capture the essence of medical expertise in everyday life. In this review, I will examine empirical evidence for and against this presumed congruence between level of socially ascribed “expertise” and performance in medicine.

Applying the Expert-Performance Approach to Medicine: Identifying Reproducibly Superior Performance on Representative Tasks in Everyday Life

The first step of the expert-performance approach involves establishing representative tasks that define the essence of the domain. The essential goals of medicine are the successful treatment of patients and the effective prevention of sickness and poor health. Medical doctors who consistently achieve the most effective treatment outcomes would, virtually by definition, be recognized as expert performers in medicine. It is, however, nearly impossible to compare the treatment success for different doctors based on their normal practice in everyday life. The treatment success of patients often depends on many factors independent of the doctors, such as the severity of the disease, the overall health of the patient, and individual differences in age, sex, and socioeconomic status. Without comparable patient characteristics, it would not be sound methodology to evaluate physicians’ expertise in treatment by directly monitoring the success of treated patients. Furthermore, it is rare that a single doctor is completely responsible for diagnosis and treatment. Most medical treatments are administered by a team of professionals. Thus, a portion of the variance in treatment outcomes is attributable to other team members and the availability of general treatment resources.

The ideal measurement situation would require that all experts and students would treat the same collection of patients. It is not feasible, however, for several doctors to treat the same patient independently. The diagnosis and treatment by the first doctor will almost invariably influence the patient and thus alter the context for subsequent doctors’ diagnoses and treatments. The traditional research method to deal with problems of this type of reactivity of

treatments is to randomly assign patients to doctors and then compare the success of the treatments. In everyday life, however, the pairing of patients to doctors is far from random. Nonetheless, even when patient populations cannot be randomly assigned, by using statistical methods to control for differences in patient variables, it is possible to estimate the treatment success of doctors holding certain credentials and characteristics. For example, Norcini et al.⁴⁷ studied which characteristics of doctors influenced the mortality and survival length of their patients with acute myocardial infarction. After statistically controlling for confounding variables, the researchers found that certification in a specialty, such as internal medicine and cardiology, led to superior performance with a lower mortality. In addition, doctors with a high volume of patients were able to reduce the length of the patients' stay in the hospital. More recently certified specialists were more able to reduce the length of treatment, which is consistent with the notion that superior knowledge, rather than effects of amount of experience per se, contributes to increased effectiveness. Analyses of millions of patients show that hospitals with a high volume of procedures have a higher treatment success than low-volume hospitals.⁴⁸ The superior results of high-volume hospitals may be due to many superior aspects of treatment, but there are demonstrations of an independent benefit for treatment when the surgery is performed by a surgeon who frequently performs that particular procedure.^{49,50} In sum, superior medical treatments appear to be linked to specialization and more extensive training in the associated domains of medicine. At present, research on superior treatment has not focused on identifying and studying individual experts who consistently perform at a superior level, but rather has searched for teams with superior learning and collaborative performance.⁵¹

Research on individual performance in medicine has primarily focused on certification and acceptable competency in medical practice and on screening students and doctors with inferior or even incorrect knowledge that might endanger their patients.⁵² Similarly, postmortem examinations have been used to assess the frequency of major errors of medical diagnosis in order to determine methods to improve medical practice and training.^{53,54} In sum, research in medical education has primarily focused on assessment of adequate and minimal competency by tests of certification. In this article, I will suggest some ways to develop a complementary approach of assessment and training that would focus on enhancing outcomes and discovering means to promote maximum levels of performance in medicine.

Can We Capture Superior Reproducible Medical Performance with Standardized Tasks?

The complex and reactive effects of medical treatment make it necessary to search for representative tasks that minimize reactivity and focus on one of the stages in the complete treatment of patients. Published research has focused on three stages that have been studied via standardized tasks, with some degree of success. The first general area of concerns the initial stage of diagnosis of pathology in perceptually available stimuli, such as x-rays, electrocardiograms (ECGs), and auscultation of heart and lung sounds. Although it is not possible to assess the correctness of the doctor's diagnosis of the patient at the original consultation in the clinic, it is possible to store the collected x-rays, ECGs, and sound recordings. For most of the patients, a valid diagnosis of the disease will eventually be found with further medical testing, surgical exploration, and development of the disease. It is possible to readminister the collected recordings of stimuli to medical doctors and students and compare their diagnoses against the correct answers. By identifying doctors with a reproducibly higher accuracy for some type of diagnosis, it is possible to study the cognitive processes mediating superior expert performance.

A second related area concerns the diagnosis of patients based on information available from a clinical interview and their medical

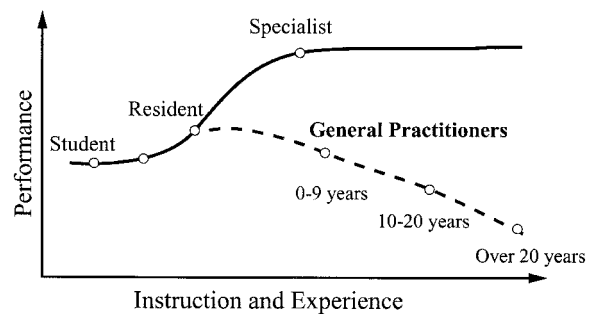


Figure 4. Two trends for development of medical performance as a function of experience and instruction.

charts. The information of actual patients can be recorded and saved until a definite diagnosis is confirmed by further medical testing and treatment. It is then possible to present a collection of these types of cases to medical doctors and students and measure the accuracy of their diagnoses, as well as monitor the cognitive processes associated with consistently superior diagnostic performance. Given that doctors nearly always see and interact with the patient, this type of test would more closely correspond to a consultation with a fellow doctor, rather than to a clinical interview in daily professional activities.

The third and final area concerns the perceptual-motor performance of treatments, such as various forms of surgery. The assumption is that one can design perceptual-motor tasks involving cadavers, simulators, or other training devices that capture the essence of specific surgical tasks involving actual patients. In sum, these three areas of medicine offer especially promising prospects where reproducibly superior performance on standardized tasks in medicine can be studied.

Perceptual diagnosis of abnormality. Treatments of medical problems require that medical professionals diagnose the underlying disease causing the associated symptoms before they prescribe effective treatments for their patients. An important stage in this diagnosis involves an examination of the patient and test results to search for abnormalities.⁵⁵ In a pioneering informal study, Butterworth and Reppert⁵⁶ presented tape recordings of heart sounds and murmurs of healthy and sick patients to many physicians and medical students. They found that the accuracy of diagnosis of the heart sounds was related to training and experience according to two different trends, which are illustrated in Figure 4. First, the diagnostic accuracy increased as a function of level of training for medical students, increased further for residents, and reached its highest observed levels for certified cardiologists. Second, doctors working in general practice did not show increased accuracy of diagnosis with further experience, and their performance instead decreased as a function of the time that had elapsed since the end of their medical training.

These two relations between different types of experience and performance have since been reported for a wide range of perceptual diagnosis skills. With respect to the first proposed relation, continued training and specialization have been shown to correspond to improved performance on standardized tests of diagnosis in several domains, with some important differences. In dermatology,⁵⁷ ECG interpretation,⁵⁸ and microscopic pathology,⁵⁹ there is a steady increase in diagnostic accuracy from around 20–40% for medical students to around 90% for experts. Interpretation of radiographs⁶⁰ increases in accuracy following a similar pattern, but the attained level of experts is somewhat lower, with accuracy of cancer detection by experts averaging around 70%.⁶¹ Studies of the reliable superiority of experts' interpretations of ECGs and radiographs show that the advantage is primarily evident for more difficult cases.^{62,63} In contrast, more recent studies of cardiac^{64,65} and pulmonary

auscultation⁶⁶ have found low rates of accuracy (20–40%) and failed to find reliable improvements among medical students and residents in the United States, though reliable but limited improvements among residents in the United Kingdom were reported.⁶⁵ The diagnostic performance of pulmonary fellows was reliably better, but still only around 60%.⁶⁶ This finding was quite lower than the accuracy estimate (80%) for cardiac auscultation among fellows and certified cardiologists reported earlier by Butterworth and Reppert.⁵⁶

With respect to the second proposed relation, research on mammographers has shown that the number of years of experience after the end of training is associated with gradual decreases in performance when factors related to training and feedback are statistically controlled.⁶⁷ Taken together, these findings support a relationship between training activities and both increased and sustained perceptual diagnostic performance, which is consistent with the importance of deliberate practice.

The effect of training with feedback for enhancing diagnostic performance is more clearly demonstrated by short-term training courses for cardiac and pulmonary auscultation.^{68–71} When students and residents were given guided practice with feedback and demonstrations of pathological conditions, their identification of abnormal heart and lung sounds improved substantially. The improvements in performance due to short-term training are not permanent, however, and reliable decreases are observed, even over a six-month retention period.⁷¹ A recent study has shown that learning to diagnose ECGs is facilitated by a deliberate-practice condition in which ECGs for different diagnoses were contrasted with a sequential presentation, compared to a more traditional approach where examples of ECGs with same diagnosis were presented together.⁷²

Analyses of the extended acquisition of diagnostic skill over years of supervised training have uncovered critical activities mediating improvement in mammography.⁵⁹ Consistent with the limited benefits derived from mere experience, improvements in performance during training are assumed to be due to the feedback that teachers give to the residents on the mammograms diagnosed in the clinic. However, during normal clinical practice malignancies are rarely encountered, so this learning mechanism is consistent with the finding that residents in mammography slowly approach the level of diagnostic accuracy of their teachers.⁵⁹ Given that the teachers' immediate diagnoses remain error prone, it will be impossible for the residents to exceed the limited accuracy level of their teachers. It is interesting to contrast that type of training situation to a hypothetical optimal scenario drawing on deliberate practice. Optimal training would allow residents to access one mammogram at a time from a large library of digitized mammograms, for which the correct diagnosis and the location of any tumors had been validated, for example, by exploratory surgery or subsequent mammograms, with the growing tumors being more clearly visible. The residents would be forced to make a detailed diagnosis of each displayed mammogram and would then receive immediate detailed feedback about the accuracy of their diagnosis, with an opportunity for reflection and reexamination of any mistakes. The library of mammograms would be indexed to encourage the resident to examine a series of related cases to facilitate detection of some critical feature or type of tumor. Learning in this type of environment should not only be much faster, but should also allow residents to reach a much higher level of diagnostic accuracy.

Recent research has explored whether it is possible for medical professionals to exceed the performance levels attained during residency by further accumulation of experience. One study involving mammogram readings found that radiologists with higher frequencies of readings were more accurate for a test sequence of mammograms than were those with a lower reading frequency.⁶¹ Subsequent research⁶⁷ has not been able to replicate the benefits of more experience, per se (a higher reading volume), because the accuracy advantage disappears when other factors are controlled

statistically. It appears that factors responsible for superior reading accuracy are related to the quality of feedback on diagnoses provided in some clinical environments. This recent controversy has motivated scientists to review carefully the current knowledge about skilled mammogram reading and call for more research clarifying the factors associated with superior diagnostic performance.⁷³

Research on the development of expert perceptual diagnosis in other domains provides suggestive insights into the necessary training conditions to reach the highest levels of performance. For example, hundreds of people have been trained to be able to identify the sex of day-old chicks with accuracies exceeding 98%.⁷⁴ Even under conditions of perfect (often immediate) feedback, it is estimated that students need to examine over 250,000 chicks in order to identify the many different patterns differentiating males and females. If these results are applicable to expert medical diagnosis in mammography, then it is possible that by increasing the frequency of targets (50% of chicks are males in chick sexing versus 1–3% malignancies for mammograms), providing improved feedback (perfect feedback versus mentor-guided feedback), and providing extended opportunities for self-study with feedback would allow experts to increase the accuracy of their medical diagnoses.

Medical diagnosis of patients. The most typical activity for doctors, especially for general practitioners, involves the diagnosis of a patient's condition during the initial encounter in the clinic. Doctors interview their patients and review their charts in order to find a diagnosis that accounts for reported symptoms. The diagnostic activity is the key to the treatment of diseases with effective standardized treatments, such as medications. In those cases, a correct diagnosis is essentially equivalent to effective treatment.

In a pioneering project to capture superior diagnostic ability among practicing physicians, Elstein et al.⁷⁵ designed high-fidelity simulations, where actors were trained to simulate patients with specific medical problems. The investigators arranged for these actors to meet doctors to obtain a diagnosis of their problems in a simulated office setting. To identify the cognitive processes mediating superior diagnostic performance, regular physicians were compared to "expert" physicians who had been nominated by their peers for their superior performance. Both types of doctors were encouraged to think aloud during their interview with the patient when there were natural breaks in the interaction between them. Surprisingly, no reliable differences in diagnostic accuracy between the peer-nominated "experts" and the regular physicians could be discovered. This lack of differentiation in diagnostic performance implies that it is not possible to identify mechanisms mediating superior diagnostic accuracy among practicing physicians—at least with Elstein et al.'s paradigm.^{76–78} Subsequent research on recertification of medical doctors has not been able to demonstrate increased accuracy of diagnosis as function of experience for doctors in general practice.⁷⁹ In fact, the performance on the certification tests decreases with age and the number of times the required recertification test has been taken.⁸⁰ Accuracy of diagnostic performance among practicing physicians is gradually reduced following graduation from medical school, in a manner similar to, albeit less pronounced than, the reduced accuracy in perceptual diagnosis illustrated in Figure 4. This decrement does not seem to reflect inevitable age-related decline, because the amount of engagement in continuing medical education activity, especially the most appropriate types of practice, has a strong relation to results on the recertification examination.⁸¹

Pioneering research on assessing diagnostic performance of medical students and residents for simulated patients has shown a striking monotonic increase in accuracy.⁸² There are several subsequent studies with different test procedures that show increases in diagnostic accuracy of students during medical school and continued supervised training^{83,84} in the same manner, though steeper, as those illustrated for perceptual diagnosis in Figure 4.

Scientists have tried to identify conditions wherein diagnostic performance systematically exceeds the stable level attained by newly certified physicians. There appear to be important differences between the diagnosis of common and rare conditions.⁸⁵ For common diseases, doctors typically attain accuracy levels at the end of the first year of residency that are comparable to levels of more experienced doctors. For rare and more complex cases in specialties, such as cardiology or nephrology, specialists are far more accurate than doctors with general experience and specialists from other areas.^{85,86} It would thus appear that the training of specialists, along with their specialized practice and teaching environment, can improve performance beyond that of a practicing physician in a manner similar to that illustrated for perceptual diagnosis in Figure 4.

In a recent review, Elstein and Schwarz⁸⁷ distinguished two types of processes that mediate medical diagnosis, which help reconcile the superior ability of specialists with the lack of individual differences among physicians in general practice. When medical conditions are frequently encountered in clinical practice, then experienced physicians will acquire patterns that allow them to recognize each condition and access mental models or prototypes for the corresponding disease. When the disease or problem is unfamiliar, however, physicians cannot draw directly on their accumulated experience and knowledge and must, therefore, rely on reasoning and systematic generation of alternatives. These two different types of diagnostic processes draw on different types of skills and organized knowledge, and their acquisition requires different types of training and deliberate practice.

During medical school and residency, there is not just an increase in accuracy of the diagnosis of common representative diseases, but there is also a change in the structure of diagnostic reasoning. With more clinical experience, biomedical reasoning during diagnosis is replaced by pattern recognition of disease schemas, which entail higher-level clinical concepts with encapsulated inferences.⁸⁸ The current focus of medical training on problem-based instruction and mentor-guided experience supports the development of these diagnostic skills and organization of knowledge to support clinical reasoning.⁸⁹ However, there might be a tradeoff between the effective acquisition of adequate proficiency, as opposed to the development of mechanisms to support continued improvements in performance following the termination of medical training. In fact, many of the effective tools for training and evaluation rely on abstracted simplified schemas, such as trained individuals simulating different diseases, which are appropriate for testing the acquisition and maintenance of proficiency. For medical professionals to be able to keep improving their diagnostic performance during years of professional practice, they would need more feedback than clinical environment naturally provides. Ideal conditions for learning would require that they get feedback, ideally immediately, on their diagnoses of actual patients in order to motivate the development of reasoning and error correction. Furthermore, they would ideally have access to a large number of patients with similar symptoms in order to develop appropriate methods for distinguishing between diseases with different treatment regimens.

Medical doctors who develop a specialization will encounter more patients with similar diseases and, thus, will have a better chance to improve their diagnostic ability, at least as long as they receive accurate informative feedback. Research on the ability to explain generated diagnoses has shown that medical specialists have a better ability to encode and accumulate relevant higher-level findings about patients, which can support specialists' reasoning about competing diagnoses.^{24,40,90}

They also display evidence of a deeper knowledge and reasoning about diseases than do less advanced doctors.⁹¹ Because there are many differences between the daily regimens of specialists and general practitioners, it will remain a challenge to identify the causal factors that can explain the development of superior diagnostic skill and reasoning within a specialty. The specialists interact with patients presenting with more similar diseases, and they typi-

cally work in clinics with both better diagnostic equipment (better feedback) and more knowledgeable colleagues. Many specialists are also actively involved in teaching and supervising students within their specialty and contribute to active research programs that shape the demands and opportunities for working specialists. Future research on expert diagnostic performers is needed to identify the types of training activities (deliberate practice) that have been most effective in helping individuals attain and maintain superior diagnostic skill in a given medical specialty. Once these activities are revealed, it should be possible to refine them further and support their development through training and continued education.

Expert performance in surgery. Many treatments in medicine are so standardized that variability in their execution and effectiveness is negligible. With treatments such as medications, there is hardly any room for individual differences in skill. In contrast, other interventions, such as surgery, are quite complex with considerable individual differences in execution. Outcomes of several surgical procedures have been found to differ as a function of the characteristics of the hospital and the surgeons treating the patients. High-volume hospitals that frequently perform a given surgical procedure or treatment have superior outcomes relative to hospitals with fewer procedures.⁹² More directly relevant to evidence for expert performance of individual surgeons, much of the variability in outcomes is related to how often a given surgeon performs the procedure, even when controlling for other variables, such as hospital volume.^{49,50} These findings clearly illustrate the benefits derived from further training and increased experience. Some investigators have, however, questioned that increased frequency of performing a surgical procedure (experience) causes superior outcome.⁹³ They raise the possibility that less able surgeons might have more initial failures, which in turn reduce their future opportunities to perform the surgical procedures. Other investigators have noticed, however, that even among surgeons with high and very high volumes of specific procedures, there were very large individual differences in outcome—far exceeding the variability that would be expected by chance factors alone.⁴⁹ The experienced surgeons with the consistently better outcomes of surgery meet the definition of expert performers.

Surgical performance is inherently interactive and highly reactive and, thus, very difficult to measure under standardized conditions. There are two types of evidence regarding its acquisition and structure, namely performance with surgical simulators and natural field experiments. A unique opportunity to study the acquisition of surgical performance under standardized conditions was offered by the introduction of new surgical techniques referred to as minimal access surgery (MAS).⁹⁴ The new skills required for MAS are fundamentally different from traditional open surgery, and the speed of acquisition does not differ between residents and experienced surgeons.⁹⁵ Furthermore, when surgical teams carry out these procedures on real patients for the first time, a marked learning effect is observed.⁹⁴ The time to complete the surgeries decreases dramatically from the first operation, and the subsequent decreases are considerable across the first ten procedures. Although the improvements associated with further surgeries diminish, the completion time is reliably reduced even after a hundred procedures. It is not just an issue of increased speed, because the frequency of injuries shows a similar decrease as a function of number of procedures performed.^{94,96}

The experience of performing a given surgical procedure with first few patients provides so much immediate natural feedback that all theories of skill acquisition predict rapid improvement of performance.² These demonstrated benefits of experience have motivated medical educators and researchers to develop simulators of surgery that would allow future surgeons to gain relevant experience before encountering their first actual patients. Those favoring this approach have argued that the designed simulators and trainers adequately simulate the task demands of the actual surgery. Early studies relied on low-tech simulators⁹⁷ and looked at transfer to

other tasks with the simulator as the criterion, such as the single interrupted stitch.^{95,97} Although these early studies were able to demonstrate transfer between performance on different tasks with the same simulator, there has been no evidence that the improvements induced by the training drills influence performance for actual surgery nor transfer to other simulators for other surgical procedures.⁹⁸ Other studies examined the effects of various abilities and other types of surgeon-specific factors^{99,100} on performance with these simulators, but found that the effects of practice on the particular simulator task tend to overshadow any effects of abilities.^{101,102}

A more empirically based method for identifying training tasks in simulators is to find benchmark tasks that capture differences in actual surgical performance between experienced surgeons, less experienced surgeons, and novices. For example, highly trained residents have been observed to perform knot-tying tasks and suturing tasks faster and with fewer movements than residents in basic surgical training.¹⁰³ Performance on these tasks can differentiate surgical residents with differing amounts of training experience.¹⁰⁴ Master surgeons made fewer errors than junior surgeons on a laparoscopic trainer,¹⁰⁵ while experienced laparoscopic surgeons performed better on the simulator than less experienced surgeons, students, and novices with respect to speed, economy, and consistency.^{106,107} A further validation of training tasks is offered by findings that experienced surgeons have superior initial performance on simulator tasks, relative to students. In addition, students trained on simulators demonstrate improved performance in actual surgeries. For example, experts with over 500 actual surgical procedures performed reliably better on a bronchoscopy simulator than two groups of doctors with less and no experience.¹⁰² When an experimental group was provided training on that simulator, they performed an actual bronchoscopy reliably better than a control group. Training on a laparoscopic simulator led to reliably fewer errors during actual surgery on an anesthetized animal.¹⁰⁸ In a recent demonstration of the value of simulator training, residents were trained to a criterion level in the laparoscopic simulator and then performed their first actual surgery with fewer errors and caused fewer injuries than did a control group of residents.¹⁰⁹ From the point of view of deliberate practice, it is important to note that the successful demonstrations of transfer from the simulator to actual surgery used explicit means to assure sufficient improvement due to simulator training. One study¹⁰⁹ required the trainees to match the simulator performance of trained surgeons. Another¹⁰² completed 20 simulations to attain a plateau, and the final study¹⁰⁸ provided the trainees with extensive opportunities to train with the simulator (ten hours).

Recent overviews and meta-analyses^{110–112} of the use and benefit of simulators in medical training have discussed the need for structured training and deliberate practice with the simulators guided by the goals of the training. Preparing surgeons for their first surgical procedure with an actual patient is probably the most important use of simulators. This training activity involves giving the trainees opportunities to perform a given a procedure while monitoring their own behavior. Unlike medical diagnosis, there are many sources of informative feedback in surgery, both in terms of immediate feedback on the execution of the procedures themselves and feedback on the health of the patient after surgery. The primary constraint for the initial acquisition of the surgery skill appears to be developing the necessary perceptual-motor coordination to carry out the procedures successfully.

Skill training of beginners is, however, only a first step toward capitalizing on the opportunities to engage in deliberate practice in potential simulators for surgery. Future simulators should allow surgeons ample opportunities to engage in deliberate practice, similar to expert performers in other domains, such as music.^{22,39,44} For example, musicians can spend hundreds of hours mastering challenging pieces in their practice room by working on selected difficult passages. Unlike surgery with actual patients, practice in

the simulator can be stopped at any time, allowing trainees an immediate chance to correct mistakes and even repeatedly perform challenging parts of procedures. It is also possible to improve the quality and detail of feedback presented to the trainees by using video recordings of the performance, both during actual and simulated procedures. Video recordings of behavior during actual surgery have been analyzed in great detail, with evaluations achieving a high level of reliability.¹¹³ With retrospective analyses of video recordings by master teachers, it would be possible to identify aspects of a surgeon's performance that can be trained and improved in the simulator—as is a common practice for enhancing the performance of athletes in soccer, football, and basketball. Finally, simulators offer the possibility of presenting rare problems and emergencies that would better prepare performers to deal with such situations. A recent study of military pilots showed that those pilots who had trained for a specific emergency situation in a simulator were more effective at responding to the same situation when it occurred during an actual flight mission.¹¹⁴ Similarly, surgeons who can experience rare emergency procedures when they are mentally ready in the simulator will be able to make necessary adjustments through additional training. These learning experiences are likely to better prepare surgeons for rare and challenging situations that occur unexpectedly. These suggestions are consistent with reviews of the extensive experience from training with flight simulators that show the importance of training designed to attain task-specific performance goals, where training is individualized to give students sufficient practice to demonstrate the required proficiency.^{115,116}

Concluding Remarks

This article has outlined how the expert-performance approach might be applied to study three types of expertise in medicine. After offering some comments on differences between acquisition of expert performance in medicine and more traditional domains of expertise, I will conclude with a general assessment of generalizable aspects of deliberate practice and how insights into the structure and acquisition in one domain of expertise can be used to guide the improvement of performance in other domains, such as medicine.

The distinctive aspect of the expert-performance approach to the study of expertise is its focus on identifying superior, reproducible behavior for representative tasks in the associated domain. The behaviors in medicine that most clearly capture the essence of expertise are effective treatments of medical conditions and diseases. Unfortunately treatments for common diseases are often highly standardized and, thus, offer only limited room for individual doctors to influence outcomes and exhibit superior performance. For example, most experienced doctors are likely equally able to diagnose and prescribe treatments for the flu and other common diseases that can be treated by prescribing familiar drugs. There are, however, many diseases that are complex and require refined diagnosis and extended treatment, such as cardiac diseases and other medical problems requiring major surgery. The activities associated with diagnosis and treatment offer considerable room for individual variability and skill that are likely to lead to differential outcomes. In this article, I identified three examples of different types of medical activities where researchers have been able to successfully identify superior performance, namely perceptual diagnosis, medical diagnosis of patients, and surgery. For each of these activities, I discussed how the investigators were able to design representative tasks that captured the essence of the medical activity and allowed standardized measurement of performance for experts and less advanced individuals. The primary focus of my discussion concerned how performance in these domains were normally acquired and how deliberate practice activities might be able to improve performance further, as well as to allow doctors to maintain their level of performance after the end of their initial medical training.

Differences between Medicine and Traditional Domains of Expertise

A successful integration of the broad body of research on expertise in a professional domain, such as medicine, requires the recognition of differences between the professional domain and more traditional domains of expertise. In contrast to competitive domains, such as sports, chess, and even music, there has been relatively little work in medicine on a "gold standard" based on measured superior performance that would allow the medical community to measure and identify superior expert levels. In medicine, experts in the medical schools and teaching hospitals frequently set the standard by which medical students and residents are evaluated and trained. Evaluation based on diagnosis of standardized patents and written tests of certification are appropriate for assessing minimal standards for acceptable proficiency in the medical profession. This type of evaluation, however, restricts the potential application of the expert-performance approach to the medical domain. Without methods for measuring reliable individual differences among the performers at the highest level, it is not possible to identify those experts with the highest objective performance and study the structure and acquisition of their performance, as is commonly done in the other domains of expertise.^{5,39,44,117} The lack of measurement procedures means that once someone has reached "expert" status, there is no accepted evaluative standard to motivate physicians to maintain and, perhaps, improve performance. Without valid informative feedback on their performance, it would be difficult for experts to engage in deliberate practice in order to enhance their skills. More critically, doctors cannot be certain of the maximal attainable level of performance for a given task unless the most able performers can demonstrate if and how such levels can be consistently achieved.

In more traditional domains of expertise, such as music and sports, demonstrations that reveal the possibility of higher performance levels have had meaningful effects on the achievements of other elite performers, as well as less accomplished individuals. Perhaps the most conspicuous example is Roger Bannister's first ever sub-four-minute mile. The earlier record for the mile run was viewed as the ultimate limit for performance, but as soon as Bannister broke the four-minute barrier, many other runners were able to do so within a couple of years.¹¹⁸ In sports, the gold medal performance at the original Olympic marathon is regularly attained by amateurs just to qualify as a participant in the Boston Marathon. Olympic swimmers from early this century would not even qualify for swim teams at competitive high-schools.²⁷ Similarly, many music pieces that were deemed unplayable are now part of the standard repertoire of students at music academies.²⁶ In each of these domains, improved methods for training and practice were largely responsible for the improved performances. In turn, future performers had access to these superior practice methods, which led to continued overall improvement. This cycle continues to enhance performance to this day.

Another difference between most professional domains, such as medicine, and the more traditional domains of expertise concerns the age of introduction and subsequent development in the domain. In medicine, most students are introduced to the field in their early 20s, whereas the introduction to training in traditional domains of expertise, such as music, ballet, chess, and sports, occurs around five to ten years of age. At the point of introduction to medical activities, most individuals are adults and have already acquired a large amount of knowledge and skills. The challenge for teachers thus becomes helping trainees not just to acquire new skills, but also to modify and adapt already acquired skills to meet the demands of medical activities. If we recognize that medical performance likely depends on preexisting knowledge and skills, we must consider the possible role of individual differences in previously acquired skills. It is possible that some aspects of these skills and knowledge structures cannot be easily modified and reorganized in a desirable manner, thus making it more difficult for some individuals to attain a high

level of performance in medicine. It is likely that as our understanding of the mechanisms mediating expert medical performance improves, it should be possible to design and refine the construction of training devices that have the dual purpose of providing opportunities for deliberate practice and assessing the current level of performance. Following the successful application of simulation technology for training pilots, it should be possible to use the simulators to provide basic training, as well as training for experienced pilots to react effectively in emergency situations.¹¹⁴ This development of training devices will make preparation in medical school and continuing education settings more individualized and effective, and will provide tools for expert performers to further enhance their levels of achievement.

Generalizable Aspects of Deliberate Practice

In this article, I have tried to show how the acquisition of superior performance in medicine is closely related to engagement in practice with feedback during medical training. I also speculated that after the end of organized medical training, continued access to conditions for deliberate practice, as well as feedback on daily medical practice, might allow doctors, especially specialists, to keep improving their performance to achieve even higher levels.

The complex integrated structure of expert performance raises many issues about how these structures can be gradually acquired and perfected over time. Whereas teachers in traditional domains of expertise start guiding the skill development of children, medical educators must encourage adult students to engage in the demanding processes involved in building the necessary skills and representations by drawing on and altering students' previously acquired skills. This type of learning is not possible without the students' full cooperation and active participation in the learning process. Students need to acquire representations that can support their planning, reasoning, and evaluation of the actual and intended performance, in order to make more appropriate adjustments to their complex skills. This advantage becomes absolutely essential at higher levels of achievement. Given that deliberate practice involves mastering tasks that students could not initially attain, or at best to attain imperfectly or unreliably, it is likely that more successful students acquire representations to support problem solving and learning through planning and analysis. Consequently, the faster learning of "talented" students might be explained by individual differences in acquired representations supporting effective learning, rather than any innate abilities or basic capacities.

From the perspective of deliberate practice, the scarcity of excellent and outstanding performance is primarily attributable to the environmental conditions necessary for its slow emergence, and to the years of deliberate practice required to develop the complex mediating mechanisms that support expertise. Even those individuals considered by themselves or others to have natural gifts gradually attain their superior performance by engaging in extended amounts of designed deliberate practice over many years. Until so-called ordinary individuals recognize that sustained effort is necessary to reach expert performance, they will continue to misattribute their inability to attain expert achievement rapidly to a lack of natural talent and will, thus, fail to reach their highest attainable level. Similarly, some people believe that decrements in performance due purely to aging are inevitable, even in the 40s, 50s, and 60s, which can again become a self-fulfilling prophecy. However, there is evidence from very demanding domains, such as music and chess, that only those older individuals who keep displaying their superior performance in public concerts and competitions are likely to be sufficiently motivated to maintain their performance by engaging regularly in appropriate deliberate practice. At present, a primary goal is to better understand the motivational factors that support and sustain continued deliberate practice in the lifelong quest for expertise in medicine and other domains.¹¹⁹

In sum, our understanding of expert medical performance will improve as we apply the expert-performance approach to the study of reproducibly superior performance in medicine. The study of expert performance in medicine is particularly likely to provide unique insights for future applications to many types of professional expertise, more generally. The domain of medicine offers a unique opportunity for research on expert performance, in that the benefits of improved medical diagnosis and treatment are quite obvious. Thus, the public is willing to invest resources to attain the highest possible levels of treatment of sickness and disease. For example, whereas some people would question the societal value of discovering training methods that would increase shot-put performance by 10%, few would dismiss the societal benefits of increasing diagnostic accuracy of mammograms by 10%. Medicine appears to provide several sufficiently constrained task domains to allow investigators in the cognitive sciences to study how a wide range of skilled professional performance areas can best be improved through appropriate training and feedback. Future research will reveal whether the promising application of the expert-performance approach to medicine will advance our understanding of expertise in this domain. This prospect is particularly exciting because in medicine, unlike sport or other competitive domains where the expert-performance approach has been most widely applied, the beneficiaries of improved performance are not only the performers themselves, but also society at large.

This work was supported in part by the FSCW/Conradi Endowment Fund of Florida State University Foundation. The author acknowledges the encouragement and continued advice from Bill McGaghie during the preparation of the RIME keynote and the writing of this article. The author also thanks many people for comments on the RIME keynote and, in particular, Geoff Norman and David Sachs for their valuable comments on earlier drafts of this article.

Correspondence: Dr. K. Anders Ericsson, Department of Psychology, Florida State University, Tallahassee, FL 32306-1270.

References

- Galton F. Hereditary Genius: An Inquiry into Its Laws and Consequences. [Originally published in 1869.] London: Julian Friedman Publishers, 1979.
- Fitts P, Posner MI. Human Performance. Belmont, CA: Brooks/Cole, 1967.
- Anderson JR. Acquisition of cognitive skill. *Psychol Rev.* 1982;89:369-406.
- Simon HA, Chase WG. Skill in chess. *Am Sci.* 1973;61:394-403.
- Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. *Psychol Rev.* 1993;100:363-406.
- Ericsson KA, Lehmann AC. Expert and exceptional performance: evidence on maximal adaptations on task constraints. *Annu Rev Psychol.* 1996;47:273-305.
- Ericsson KA. The scientific study of expert levels of performance: general implications for optimal learning and creativity. *High Abil Stud.* 1998;9:75-100.
- Doane SM, Pellegrino JW, Klatzky RL. Expertise in a computer operating system: conceptualization and performance. *Hum Comput Interact.* 1990;5:267-304.
- Reif F, Allen S. Cognition for interpreting scientific concepts: a study of acceleration. *Cogn Instruct.* 1992;9:1-44.
- Dawes RM. House of Cards: Psychology and Psychotherapy Built on Myth. New York: Free Press, 1994.
- Rosson MB. The role of experience in editing. In: Proceedings of INTERACT '84 IFIP Conference on Hum-Comput Interact. New York: Elsevier, 1985:45-50.
- Ashworth CA. Skill as the fit between performer resources and task demands. In: Proceedings of 14th Annual Cognitive Science Meeting. Hillsdale, NJ: Erlbaum, 1992:444-9.
- Gawel R. The use of language by trained and untrained experienced wine tasters. *J Sens Stud.* 1997;12:267-84.
- Valentin D, Pichon M, de Boishebert, Abdi H. What's in a wine name? When and why do wine experts perform better than novices? *Abstr Psychonom Soc.* 2000;5:36.
- Camerer CF, Johnson EJ. The process-performance paradox in expert judgment: how can the experts know so much and predict so badly? In: Ericsson KA, Smith J (eds). *Towards a General Theory of Expertise: Prospects and Limits.* Cambridge: Cambridge University Press, 1991:195-217.
- Shanteau J, Stewart TR. Why study expert decision making? Some historical perspectives and comments. *Organ Behav Hum Decis.* 1992;53:95-106.
- Howe MJA, Davidson JW, Sloboda JA. Innate talents: reality or myth? *Behav Brain Sci.* 1998;21:399-442.
- Ericsson KA. Basic capacities can be modified or circumvented by deliberate practice: a rejection of talent accounts of expert performance. A commentary on M. J. A. Howe, J. W. Davidson, and J. A. Sloboda "Innate Talents: Reality or Myth?" *Behav Brain Sci.* 1998;21:413-4.
- Ericsson KA. Peak performance and age: an examination of peak performance in sports. In: Baltes PB, Baltes MM (eds). *Successful Aging: Perspectives from the Behavioral Sciences.* New York: Cambridge University Press, 1990:164-95.
- Ericsson KA. The acquisition of expert performance: an introduction to some of the issues. In: Ericsson KA (ed). *The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports, and Games.* Mahwah, NJ: Erlbaum, 1996:1-50.
- Ericsson KA, Smith J. Prospects and limits in the empirical study of expertise: an introduction. In: Ericsson KA, Smith J (eds). *Toward a General Theory of Expertise: Prospects and Limits.* Cambridge: Cambridge University Press, 1991:1-38.
- Ericsson KA. The path to expert performance: insights from the masters on how to improve performance by deliberate practice. In: Thomas P (ed). *Optimizing Performance in Golf.* Brisbane: Australian Academic Press, 2001:1-57.
- de Groot A. Thought and Choice in Chess [Original work published 1946.] The Hague: Mouton, 1978.
- Ericsson KA, Patel VL, Kintsch W. How experts' adaptations to representative task demands account for the expertise effect in memory recall: comment on Vicente and Wang (1998). *Psychol Rev.* 2000;107:578-92.
- Bloom BS. Generalizations about talent development. In: Bloom BS (ed). *Developing Talent in Young People.* New York: Ballantine Books, 1985:507-49.
- Lehmann AC, Ericsson KA. The historical development of domains of expertise: performance standards and innovations in music. In: Steptoe A (ed). *Genius and the Mind.* Oxford: Oxford University Press, 1998:67-94.
- Schulz R, Curnow C. Peak performance and age among superathletes: track and field, swimming, baseball, tennis, and golf. *J Gerontol Psychol Sci.* 1988;43:113-20.
- Krampe RT, Ericsson KA. Maintaining excellence: deliberate practice and elite performance in young and older pianists. *J Exp Psychol Gen.* 1996;125:331-59.
- Charness N, Krampe RT, Mayr U. The role of practice and coaching in entrepreneurial skill domains: an international comparison of life-span chess skill acquisition. In: Ericsson KA (ed). *The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports, and Games.* Mahwah, NJ: Erlbaum, 1996:51-80.
- Lehmann AC, Ericsson KA. Music performance without preparation: structure and acquisition of expert sight-reading. *Psychomusicology.* 1996;15:1-29.
- Sloboda JA. The acquisition of musical performance expertise: deconstructing the "talent" account of individual differences in musical expressivity. In: Ericsson KA (ed). *The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports, and Games.* Mahwah, NJ: Erlbaum, 1996:107-26.
- Starkes JL, Deakin J, Allard F, Hodges NJ, Hayes A. Deliberate practice in sports: what is it anyway? In: Ericsson KA (ed). *The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports, and Games.* Mahwah, NJ: Erlbaum, 1996:81-106.
- Helsen WF, Starkes JL, Hodges NJ. Team sports and the theory of deliberate practice. *J Sport Exerc Psychol.* 1998;20:12-34.
- Ward P, Hodges NJ, Williams AM, Starkes JL. Deliberate practice and expert performance. In: Williams AM, Hodges NJ (eds). *Skill Acquisition in Sport.* London: Routledge, 2004:231-58.
- Schneider W. Acquiring expertise: determinants of exceptional performance. In: Heller KA, Mönks J, Passow H (eds). *International Handbook of Research and Development of Giftedness and Talent.* Oxford: Pergamon Press, 1993:311-24.
- Masunaga H, Horn J. Expertise and age-related changes in components of intelligence. *Psychol Aging.* 2001;16:293-311.
- Starkes JL, Weir PL, Young BW. Retaining expertise: what does it take for older expert athletes to continue to excel? In: Starkes J, Ericsson KA (eds). *Expert Performance in Sport: Recent Advances in Research on Sport Expertise.* Champaign, IL: Human Kinetics, 2003:275-93.
- Ericsson KA. How experts attain and maintain superior performance: implications for the enhancement of skilled performance in older individuals. *J Aging Phys Activ.* 2000;8:346-52.
- Ericsson KA. Attaining excellence through deliberate practice: Insights from the study of expert performance. In: Ferrari M (ed). *The Pursuit of Excellence in Education.* Hillsdale, NJ: Erlbaum, 2002:21-55.
- Ericsson KA, Kintsch W. Long-term working memory. *Psychol Rev.* 1995;102:211-45.
- Salthouse TA. Effects of age and skill in typing. *J Exp Psychol Gen.* 1984;113:345-71.
- Abernethy B. Visual search strategies and decision-making in sport. *Int J Sports Psychol.* 1991;22:189-210.
- Williams AM, Ward P, Knowles JM, Smeeton NJ. Anticipation skill in a real-world task: measurement, training, and transfer in tennis. *J Exp Psychol Appl.* 2002;8:259-70.
- Ericsson KA. The development of elite performance and deliberate practice: an update from the perspective of the expert-performance approach. In: Starkes J, Ericsson KA (eds). *Expert Performance in Sport: Recent Advances in Research on Sport Expertise.* Champaign, IL: Human Kinetics, 2003:49-81.
- Ericsson KA. How the expert-performance approach differs from traditional approaches to expertise in sports: in search of a shared theoretical framework for studying expert performance. In: Starkes J, Ericsson KA (eds). *Expert Performance in Sport: Recent Advances in Research on Sport Expertise.* Champaign, IL: Human Kinetics, 2003:371-401.

46. Davidson PM. The surgeon for the future and implications for training. *Aust N Z J Surg.* 2002;72:822-8.
47. Norcini JJ, Lipner RS, Kimball HR. Certifying examination performance and patient outcomes following acute myocardial infarction. *Med Educ.* 2002;36:853-9.
48. Birkmeyer JD, Siewers AE, Finlayson EV, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med.* 2002;346:1128-37.
49. Begg CB, Riedel ER, Bach PB, et al. Variations in morbidity after radical prostatectomy. *N Engl J Med.* 2002;346:1138-44.
50. Schrag D, Panageas KS, Riedel E, et al. Hospital and surgeon procedure volume as proctors of outcome following rectal cancer resection. *Ann Surg.* 2002;236:583-92.
51. Pisano GP, Bohmer RMJ, Edmondson AC. Organizational differences in rates of learning: evidence from the adoption of minimally invasive cardiac surgery. *Manage Sci.* 2001;47:752-68.
52. Hutchinson L, Aiken P, Hayes T. Are medical postgraduate certification processes valid? A systematic review of the published evidence. *Med Educ.* 2002;36:73-91.
53. Bordage G. Why did I miss the diagnosis? Some cognitive explanations and educational implications. *Acad Med.* 1999;74 suppl 10:S138-43.
54. Sonderegger-Iseli K, Burger S, Muntwyler J, Salomon F. Diagnostic errors in three medical areas: a necropsy study. *Lancet.* 2000;355:2027-31.
55. Norman GR, Coblenz CL, Brooks LR, Babcock CJ. Expertise in visual diagnosis: a review of the literature. *Acad Med.* 1992;69(suppl 10):S78-83.
56. Butterworth JS, Reppert EH. Auscultatory acumen in the general medical population. *JAMA.* 1960;174:32-4.
57. Norman GR, Rosenthal D, Brooks LR, Allen SW, Muzzin LJ. The development of expertise in dermatology. *Arch Dermatol.* 1989;125:1063-8.
58. Hatala R, Norman GR, Brooks LR. Impact of a clinical scenario on accuracy of electrocardiogram interpretation. *J Gen Intern Med.* 1999;14:126-9.
59. Crowley RS, Naus GJ, Stewart J, Friedman CP. Development of visual diagnostic expertise in pathology: an information-processing study. *J Am Med Inform Assoc.* 2003;10:39-51.
60. Nodine CF, Kundel HL, Mello-Thoms C, et al. How experience and training influence mammography expertise. *Acad Radiol.* 1999;6:575-85.
61. Esserman L, Cowley H, Eberle C, et al. Improving the accuracy of mammography: volume and outcome relationships. *J Natl Cancer Inst.* 2002;94:369-75.
62. Simpson SA, Gilhooly KJ. Diagnostic thinking processes: evidence from a constructive interaction study of electrocardiogram (ECG) interpretation. *Appl Cogn Psychol.* 1997;11:543-54.
63. Wolf FM, Miller JG, Bozynski ME, et al. Effects of experience and case difficulty on the interpretation of pediatric radiographs. *Acad Med.* 1994;69(suppl 10):S31-3.
64. Mangione S, Nieman LZ. Cardiac auscultatory skills of internal medicine and family practice trainees: a comparison of diagnostic proficiency. *JAMA.* 1997;278:717-22.
65. Mangione S. Cardiac auscultatory skills of physicians-in-training: a comparison of three English-speaking countries. *Am J Med.* 2003;110:210-6.
66. Mangione S, Nieman LZ. Pulmonary auscultatory skills during training in internal medicine and family practice. *Am J Respir Crit Care Med.* 1999;159:1119-24.
67. Beam CA, Conant EF, Sickles EA. Association of volume and volume-independent factors with accuracy in screening mammogram interpretation. *J Natl Cancer Inst.* 2003;95:282-90.
68. Horiszny JA. Teaching cardiac auscultation using simulated heart sounds and small group discussion. *Fam Med.* 2001;33:39-44.
69. Issenberg SB, Petrusa ER, McGaghie WC, et al. Effectiveness of a computer-based system to teach bedside cardiology. *Acad Med.* 1999;74(suppl 10):S93-5.
70. Issenberg SB, McGaghie WC, Gordon DL, et al. Effectiveness of a cardiology deliver course for internal medicine residents using simulation technology and deliberate practice. *Teach Learn Med.* 2002;14:223-8.
71. McGuire C, Hurley RE, Babbott D, Butterworth JS. Auscultatory skill: gain and retention after intensive instruction. *J Med Educ.* 1964;39:120-30.
72. Hatala RM, Brooks LR, Norman GR. Practice makes perfect: the critical role of mixed practice in the acquisition of ECG interpretation skills. *Adv Health Sci Educ.* 2003;8:17-26.
73. Elmore JG, Miglioretti DL, Carney PA. Does practice make perfect when interpreting mammography? Part II. *J Natl Cancer Inst.* 2003;95:250-2.
74. Martin RD. *The Specialist Chick Sexer: A History—Worldview—Future Prospects.* Box Hill South, Australia: Bernal Publishing, 1994.
75. Elstein AS, Shulman LS, Sprafka SA. *Medical Problem Solving: An Analysis of Clinical Reasoning.* Cambridge, MA: Harvard University Press, 1978.
76. Elstein AS, Shulman LS, Sprafka SA. Medical problem solving: a ten-year retrospective. *Eval Health Prof.* 1990;13:5-36.
77. McGaghie WC. Medical problem-solving: a reanalysis. *J Med Educ.* 1980;55:912-21.
78. Patel VL, Groen GJ. Knowledge based solution strategies in medical reasoning. *Cogn Sci.* 1989;10:91-116.
79. Eva KW. The aging physician: changes in cognitive processing and their impact on medical practice. *Acad Med.* 2002;77(suppl 10):S1-6.
80. Leigh TM, Young PR, Haley JV. Performance of family practice diplomats on successive mandatory recertification examinations. *Acad Med.* 1993;68:912-8.
81. Rhodes RS, Biesten TW, Ritchie WP, Malangoni MA. Continuing medical education activity and American Board of Surgery Examination performance. *J Am Coll Surgeons.* 2003;196:604-9.
82. Neufeld VR, Norman GR, Feightner JW, Barrows HS. Clinical problem-solving by medical students: a longitudinal and cross-sectional analysis. *Med Educ.* 1981;15:315-22.
83. Kramer AW, Dusman H, Tan LH, Jansen KJ, Grol RP, van der Vleuten CP. Effect of extension of postgraduate training in general practice on the acquisition of knowledge of trainees. *Fam Pract.* 2003;20:207-12.
84. Van de Wiel, MWJ, Boshuizen HPA, Schmidt HG. Knowledge restructuring in expertise development: evidence from pathophysiological representations of clinical cases by students and physicians. *Eur J Cogn Psychol.* 2000;12:323-55.
85. Norman GR, Trott AD, Brooks LR, Smith EKM. Cognitive differences in clinical reasoning related to postgraduate training. *Teach Learn Med.* 1994;6:114-20.
86. Rikers RMJP, Schmidt HG, Boshuizen HPA. On the constraints of encapsulated knowledge: clinical case representations by medical experts and subexperts. *Cogn Instruct.* 2002;20:27-45.
87. Elstein AS, Schwarz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ.* 2002;324:729-32.
88. Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise: theory and implications. *Acad Med.* 1990;65:611-21.
89. Norman GR. Research in medical education: three decades of progress. *BMJ.* 2002;324:1560-2.
90. Patel VL, Arocha JF, Kaufmann DR. Diagnostic reasoning and medical expertise. *Psychol Learn Motiv.* 1994;30:187-251.
91. Lemieux M, Bordage G. Propositional versus structural semantic analyses of medical diagnostic thinking. *Cogn Sci.* 1992;16:185-204.
92. Hillner BE, Smith TJ, Desch CE. Hospital and physician volume or specialization and outcomes in cancer treatment: importance in quality of cancer care. *J Clin Oncol.* 2000;18:2327-40.
93. Cowan JA, Dimick JB, Thompson BG, Stanley JC, Upchurch GR. Surgeon volume as an indicator of outcomes after carotid endarterectomy: an effect independent of specialty practice and hospital volume. *J Am Coll Surgeons.* 2002;195:814-21.
94. Rogers DA, Elstein AS, Bordage G. Improving continuing medical education for surgical techniques: applying the lessons learned in the first decade of minimal access surgery. *Ann Surg.* 2001;233:159-66.
95. Rosser JC Jr, Rosser LE, Savalgi RS. Objective evaluation of a laparoscopic surgical skill program for residents and senior surgeons. *Arch Surg.* 1998;133:657-61.
96. Moore MJ, Bennett CL. The learning curve for laparoscopic cholecystectomy. The Southern Surgeons Club. *Am J Surg.* 1995;170:55-9.
97. Rosser JC, Rosser LE, Savalgi RS. Skill acquisition and assessment of laparoscopic surgery. *Arch Surg.* 1997;132:200-4.
98. Ström P, Kjellin A, Hedman L, Wredmark T, Felländer-Tsai L. Training in tasks with different visual-spatial components does not improve virtual arthroscopy performance. *Surg Endosc.* 2004;18:115-20.
99. Risucci D, Geiss A, Gellman L, Pinard B, Rosser J. Surgeon-specific factors in the acquisition of laparoscopic surgical skills. *Am J Surg.* 2001;181:289-93.
100. Wanzel KR, Hamstra SJ, Anastakis DJ, Matsumoto ED, Cusimano MD. Effects of visual-spatial ability on learning spatially-complex surgical skills. *Lancet.* 2002;359:230-1.
101. MacDonald J, Willimas RG, Rogers DA. Self-assessment in simulation-based surgical skills training. *Am J Surg.* 2003;185:319-22.
102. Ost D, DeRosiers A, Britt EJ, Fein AM, Lesser ML, Mehta AC. Assessment of a bronchoscopy simulator. *Am J Respir Crit Care Med.* 2001;164:2248-55.
103. Bann SD, Khan MS, Darzi AW. Measurement of surgical dexterity using motion analysis of simple bench tasks. *World J Surg.* 2003;27:390-4.
104. Bann SD, Kwok KF, Lo CY, Darzi A, Wong J. Objective assessment of technical skills of surgical trainees in Hong Kong. *Br J Surg.* 2003;90:1294-9.
105. Francis NK, Hanna GB, Cuschieri A. The performance of master surgeons on the advanced Dundee endoscopic psychometric tester: contrast validity study. *Arch Surg.* 2002;137:841-4.
106. Adrales GL, Park AE, Chu UB, et al. A valid method of laparoscopic simulation training and competence assessment. *J Surg Res.* 2003;114:156-62.
107. Gallagher AG, Satava RM. Virtual reality as a metric for the assessment of laparoscopic psychomotor skills. *Surg Endosc.* 1967;16:1746-52.
108. Hyltander A, Liljegren E, Rhodin PH, Lönröth H. The transfer of basic skills learned in a laparoscopic simulator to the operating room. *Surg Endosc.* 2002;16:1324-8.
109. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance. *Ann Surg.* 2002;236:458-64.
110. Kneebone R. Simulation in surgical training: educational issues and practical applications. *Med Educ.* 2003;37:267-77.
111. Issenberg B. Simulation—the evidence. Presented at the Simulation in Medical Education conference. London: The Association for the Study of Medical Education, 2004.
112. Reznick M, Harter P, Krummel T. Virtual reality and simulation: training the future emergency physician. *Acad Emerg Med.* 2002;9:78-87.
113. Eubanks TR, Clements RH, Pohl D, et al. An objective scoring system for laparoscopic cholecystectomy. *J Am Coll Surg.* 1999;189:566-74.
114. McKinney EH, Davis KJ. Effects of deliberate practice on crisis decision performance. *Hum Factors.* 2003;45:436-44.
115. Bell HH, Waag WL. Evaluating the effectiveness of flight simulators for training combat skills: a review. *Int J Aviat Psychol.* 1998;8:223-42.
116. Hays RT, Jacobs JW, Prince C, Salas E. Flight simulator training effectiveness: a meta analysis. *Mil Psychol.* 1992;4:63-74.
117. Ericsson KA, Charness N. Expert performance: its structure and acquisition. *Am Psychol.* 1994;49:725-47.
118. Denison J. *Bannister and Beyond: The Mystique of the Four Minute Mile.* Halcottsville, NY: Breakaway Books, 2003.
119. Guest CB, Regehr G, Tiberius RG. The lifelong challenge of expertise. *Med Educ.* 2001;35:78-81.