Competency Based Medical Education is Wrong for Surgery

Thane A. Blinman, MD, FACS

¹Abstract—Competency Based Medical Education CBME has been pushed into place as the dominant paradigm for medical and surgical education. Premised on the notion that medical proficiency can be rendered into predefined observable behaviors, CBME takes an explicitly reductionistic and behaviorist approach to professional education. Unfortunately, this approach is fatally flawed, relying on false premises of human action, employing unworkable methods, and unsupported by empirical data. Deployment of CBME has produced disintegration of surgical training, with loss of technical ability, degradation of professional ownership, and worsening clinical judgment. In contrast, an approach that updates working Halstedian techniques, explicitly favors cognitive integration, and deemphasizes artificial metrics of advancement holds more promise for creating proficient surgeons. (Preprint, Jan 2017)

Index Terms—medical education; competencies; behaviorism; expertise; deliberate practice; concept theory

I. INTRODUCTION

In 1997, Cameron delivered a discussion of William Halsted, citing Halsted's invention 100 years earlier of the surgical residency: "It is this method for training surgeons introduced by Halsted that probably is more responsible than any other single factor for the incredible productivity that has placed the United States in the forefront of surgical science throughout the world" (1). Halsted's advance was placing the chief resident rather than the professor at the center of a training program. Halsted, and those who emulated him, expected that energetic self-education through experience on the wards would rival or eclipse learning delivered by the professor. Halsted's model worked. In the ensuing decades, general surgeons became remarkably proficient in advancing and practicing surgical techniques that brought unprecedented comfort and cure to countless patients. Surgery in the century after Halsted was radically better than surgery in the century before, and Halsted's approach to the education of surgeons was the essential vehicle of those advances. The methods initiated by Halsted and imitated and modified by his first students have stood with good reason as revered surgical tradition.

But tradition had limits. Despite the demonstrable successes of the profession, leaders in medical education had

been scrutinizing the methods and results of the Halstedian model, and had found serious shortcomings. The old system was flawed, they observed, even dangerous, abusing trainees and imperiling patients (2). As these opinions spread, reform of medical education became a moral imperative.

An ironic truth of surgical education is that there is hardly any training in training programs. Since Halsted, the lineaments of surgical education have relied upon intense and lengthy bedside experiences and a method of learn-by-doing at the side of experienced surgeons. Rather than some closed curriculum that methodically and repetitively drills a structured body of knowledge or a delimited set of skills, residency training has instead relied upon an open scope and a long soak in practical experience. Surgical residents are deliberately selected to exhibit a hungry curiosity and forceful autodidactic capability. Halstedian residency extends methods from classic apprenticeship, where one learns by watching and imitating, gradually being allowed to venture into tasks that require greater ability (and that have higher stakes). Humans have mastered professions this way for millennia.

But the traditional model of surgical residency had problems. For example, it relied too much on time, long hours over many years, to confer expertise. Eventually, these long hours were seen as unreasonable, burdensome, and even unsafe (3, 4). When the ACGME mandated that residents' hours be limited, experiential time was curtailed, leaving residents with at least 20% less clinical exposure (5). However, nothing replaced this experiential exposure. In other words, in the name of safety, the ACGME reduced the single most important mechanism by which new surgeons learn, the means by which any expert achieves expertise, "filling the experience bank" (6) over extended time and practice (7). The result was diminished and diluted clinical exposure in residency, necessarily conferring less experience, and therefore less expertise, to each trainee.

But this time-dependency was not the only weakness of the Halstedian surgical residency model. Apprenticeship type training could be inherently conservative and slow to incorporate new advances. For example, most progress in minimally invasive methods in academic hospitals were not achieved by established physicians pushing, but by young trainees pulling. Adoption of minimally-invasive methods in the 1990's was often driven by trainees eager to include these new methods into their experience. At the same time, these efforts were resisted by some established surgeons, suspicious of these strange new methods. Minimally invasive methods gained currency not by widespread adoption by established surgeons, but through entry into practice by trainees who had learned from a few mavericks (or who had taught themselves).

1. T. Blinman is associate professor of surgery and training program director for the Division of General, Thoracic and Fetal Surgery, at the Children's Hospital of Philadelphia, 3401 Civic Center Blvd, Philadelphia, PA, 19104 (e-mail: blinman@email.chop.edu).

Whether such conservative tendency in surgery is more feature or bug is an interesting topic, but when trainees represent the vanguard of clinical practice, safeguards for patients may be uncertain.

More importantly, traditional residency relies heavily on imitation and inference, but inferential learning can be uneven. For example, residency immersion leaves trainees prone to propagation of surgical myths (8, 9), that while untrue, nevertheless resist eradication because they are established in a culture of common practice. More commonly, holes in residents' knowledge are neither recognized nor filled. For example, few residents master the oxygen delivery equation for critical care (10), or the principles of electrosurgery (11), or explicit strategies for maximizing precision and mechanical advantage in minimally-invasive methods (12). Young learners begin to resuscitate and operate, but their actions are determined in substantial part, not by "formal" instruction, but by imitation, by aping others. The nascent surgeon, when presented with a clinical situation that demands action, does not remember the last lecture, but the last experience: *he does what he has seen done*. For humans, imitation is an essential mechanism of learning (13). But this route to knowledge is also prone to cargo-cult type errors, where actions are imitated while the underlying rationale or scientific basis for some practice (and, critically, its exceptions) remain unstated, misunderstood, or misapplied. Such an unsure method could not form the basis of modern, accredited, professional education.

Finally, accreditors believed that the traditional residency was insufficiently legible in terms of content, controls, and measurements. Aside from the knowledge testing and case-type exposure required by the American Board of Surgery (ABS), no other objective outcomes allowed programs to demonstrate the effectiveness of their teaching or of their graduates' broad clinical competence. There was little accountability, and no one could see the broad state of professional training. Doubts even arose about the value of ABS certification as a mark of surgical competence (14). The graduates might be good, but who could say with any veracity?

For all these reasons, leaders in GME decided a new way of training doctors was required, a method that they believed would allow assessment of learners' progress based on predefined outcomes that could be written down and observed. This new paradigm came to be called Competency-Based Medical Education (CBME).

The Accreditation Council for Graduate Medical Education (ACGME) has driven notable changes to residency structure and resident assessment, and all of these changes are instances of CBME in application. Since as early as 1999, CBME has supplied the theory behind virtually every new accreditation mandate, and has stood as the default rubric for organizing all accredited medical and surgical training programs. Across the United States, training programs have dutifully complied with ACGME mandates, incorporating CBME methods into their curricula and assessments. Educational authorities have enthusiastically encouraged these changes and praised the visionary new medical education model.

There is only one problem: CBME doesn't work.

The purpose of this review is to refute the tenets of CBME, and show the incompatibility of this pedagogical theory with surgical education. First, the historical rise of CBME is briefly outlined. Next, CBME, as manifested in the ACGME's Core Competencies and Milestones, is revealed to be unsupported by empirical evaluation and premised on an obsolete model of human action. When this premise is tested in actual application, it is shown to lead to a cascade of failures that seriously damage surgical education. Fortunately, where CBME fails to deliver improved surgical education, other methods can be found that offer the objectivity and transparency promised (but not delivered) by CBME. Alternatively, surgical training based on cognitively integrated concepts of human proficiency may preserve effective elements of traditional Halstedian training while circumventing its shortfalls.

II. THE RISE OF COMPETENCY BASED MEDICAL EDUCATION

Competency-based medical education has become the dominant model of medical education since the introduction of the "core competencies" in 1999. Driven by imperatives from the ACGME, graduate medical education (GME) has been relentlessly reshaped according to competency-based methods, with changes spurred by regular advocacy. The ACGME offers frequent educational courses to familiarize program directors and coordinators in competency-based doctrines, and to aid programs in maintaining compliance. Papers in the ACGME's publication, the Journal of Graduate Medical Education, regularly offer advice for implementing and extending competency-based assessments (e.g. (14), (15), (16), (17)). A group called the International Competency-Based Medical Education Collaborators, forcefully argues for CBME, aiming to "accelerate its uptake worldwide" (18). Virtually every conference on medical education begs the competency-based question, discussions devoted to methods of implementing competency oriented methods (if not to testing their effectiveness). The proponents for CBME have an agenda (19), the pursuit of which has been deliberate, relentless, and nearly unopposed.

CBME is often described in counterpoint to what it aims to supplant, "time-based" education. Advocates of CBME see residency training comprised by a defined training time (e.g. a 5-year residency), an abstractly defined operative experience, and a broadly described "learning environment" as merely a form of process control, divorced from actual outcomes. In this way, traditional residency training is regarded as insufficiently legible, measureable, or effective. In contrast, "Harden and colleagues wrote that [CBME] was a visionary new approach to medical education because it emphasized the type of doctor produced rather than the process of education (20)." Proponents of CBME assert that competencies could define "the good doctor" in a granular, discrete, transparent, and most importantly, measurable way.

Although notions of CBME, also sometimes called "objectives based education" (OBE), have been discussed for decades, authors struggle to define CBME, admitting it "is easy to conceptualize, but difficult to define" (20). Frank, et al (21) provide an exhaustive survey of CBME literature,

isolating its salient feature (emphasis added): “Nearly every publication emphasized the cardinal idea that CBME is a distinct approach because of its dedication to *predefined graduate abilities as the organizing principle*.” After reviewing dozens of papers, the authors settle on the following definition of CBME:

“Competency-based medical education (CBME) is an approach to preparing physicians for practice that is fundamentally oriented to graduate outcome abilities and organized around competencies derived from an analysis of societal and patient needs. It de-emphasizes time-based training and promises greater accountability, flexibility, and learner-centeredness.”

This definition is long on promises (“accountability, flexibility, learner-centeredness”) but short on methods. Nevertheless, CBME explicitly relies on highly granular descriptions of particular, narrowly-defined skills and knowledge, also called “competencies” or sometimes “entrustable professional activities” (EPAs), or even “statements of awarded responsibility” (STARS). (22). Explicitly reductionist, CBME “...invites deconstructing a given competency into smaller parts and can be done using the smallest part as representative of the larger competency without compromising the larger interspecialty conversations that have proven so useful” (23). Perhaps one might call a competency an atom of professional ability.

When taken in aggregate, competencies should comprise an able physician of any specialty. CBME is premised on the ideas that one can define specific elements of “the good doctor” (20); that most of these elements are the same across specialties (24) (25); and that demonstration by the learner of mastery of these elements provides an objective measure of a trainee’s progress along a predefined gradient toward this archetype (26). “Outcome-based frameworks require a defined scheme of levels of progression towards the outcome.” (27) The “outcomes” here are “observable behaviors” exhibited by the trainee. Today, these outcomes are encoded by the ACGME as the “Milestones”.

The Milestones, as defined by the ACGME (<http://www.acgme.org/What-We-Do/Accreditation/Milestones/Overview>), emphasize this training gradient, promising measurable “progression toward the outcome.” Milestones are “competency-based developmental outcomes (e.g., knowledge, skills, attitudes, and performance) that can be demonstrated progressively by residents and fellows from the beginning of their education through graduation to the unsupervised practice of their specialties.” Incorporating CBME through the Milestones is promised to bring many benefits, including “continuous monitoring”, “public accountability”, “explicit and transparent expectations of performance,” and several other desiderata. In practice, the Milestones are specialty-specific lists of more or less abstracted behaviors that can be assessed according to preset expectations that correspond to each learner’s time in the program. To hasten adoption of this scheme, the ACGME mandates that all accredited medical training programs report aggregated scores for each resident semi-annually.

Several of the competencies are regarded as being so fundamental and universal that they apply to any physician, in any specialty. These “core competencies” (CC) include six facets of physician cognitive makeup and performance that a panel of experts designated as the essential elements of any able physician (Table 1). The core competencies bear the force of the authority of the ACGME, and assessment and reporting of these now form the basis of physician training and evaluation from medical school through professional certification.

III. FAILURES OF COMPETENCY BASED MEDICAL EDUCATION

CBME seems at first to be reasonable, responsible, and practical. But even advocates allow that these impressions do not survive attempts to actually implement these ideas (24). As instantiated as the “Milestones”, CBME displays critical failures that preclude its promised benefits. This section traces how CBME marks a chain of failures from conceptual premises to practical implementation.

A. Bad Ideas: Conceptual Disintegration

The labored definition of CBME hints at the root problem with CBME as an idea. Philosophically, CBME represents a product of cognitive conceptual disintegration (28). In particular, CBME fails to identify discriminating features of its concepts of medical proficiency, or to define these features coherently. What features it does identify, it fails to assemble logically into cognitively related wholes, choosing instead to enumerate lists of unconnected, or disintegrated “behaviors.” This tendency toward scattered ideas can be seen in the original competencies invoked in the original Outcomes Project, the core competencies.

Despite more than a decade of official advocacy and mandates, few trainees can recite the CC (29), and virtually no one in active clinical practice understands most of these ideas. Because the CC lack face validity (30), the assertion that these six descriptors are somehow measurable parameters is not supported by parsing their language (TABLE 1). For this reason, CC are often mocked as “the rogue competencies,” resistant to usable interpretation.

Aside from the ideas of “patient care” and “medical knowledge”, the CC are hobbled by internal contradictions. For example, “systems-based practice” is probably intended to acknowledge that physicians always work within constraints imposed by the system (not defined) in which they work. But fuzzy terms like “responsiveness” confuse anyone trying to “demonstrate” awareness of the system. The language seems to call for collegiality and team work or some other form of egalitarian professional manners. However, the description continues, seeming to blur the distinction between individual patients’ needs, and the population-based uniformity of guidelines, pathways, and other systemic efforts intended to reduce practice variation. This kind of “systems-based practice” creates a moral hazard for a surgeon trying to find the best intervention for an individual patient. Instead, he must “demonstrate responsiveness” by prioritizing the “larger context” of the health care system. “Larger context” is code

for the group, the average, the population, or society. Mis-prioritizing collective metrics over individual results has yielded many modern medical absurdities, such as the mandatory data collection of “meaningful use” (“Do you smoke?” “No, I’m only six!”), context-dropping clinical pathways, and other well-meaning but dangerous clinical corruptions. “Systems-based practice” represents either a banal tautology or an immoral call for the “system” to enjoy priority when making clinical decisions for individuals. This incoherence is a result of trying to capture concepts of medical ethics according to a positivist and reductionist methodology (31).

The other CC are similarly incoherent. “Practice based learning” seems to call for a kind of meta-cognition, the kind of personal and introspective autodidactic learning that is essential to attaining proficiency, but impossible to measure using the Milestones’ vague language and blunt scale. “Professionalism” does not provide a means to discriminate between opposite professional actions (e.g. “the surgeon rushed the patient to the operating room”; “the surgeon decided operating might be too dangerous”), either of which could demonstrate a “commitment to carrying out professional activities” (another tautology) or abnegation of such commitment. “Interpersonal and communication skills” hinges upon “effective information exchange”, as if the resident was merely some medium, and introduces the new-age verb “teaming,” blurring this competency with “systems based practice.” And so on.

TABLE I
THE CORE COMPETENCIES AS DEFINED BY THE ACGME

Patient Care	Residents must be able to provide patient care that is compassionate, appropriate, and effective for the treatment of health problems and the promotion of health.
Medical Knowledge	Residents must be able to demonstrate knowledge about established and evolving biomedical, clinical, and cognate (e.g. epidemiological and social-behavioral) sciences and the application of this knowledge to patient care.
Practice-Based Learning and Improvement	Residents must be able to investigate and evaluate their patient care practices, appraise and assimilate scientific evidence, and improve their patient care practices.
Interpersonal and Communication Skills	Residents must be able to demonstrate interpersonal and communication skills that result in effective information exchange and teaming with patients, patients’ families, and professional associates.
Professionalism	Residents must be able to demonstrate a commitment to carrying out professional responsibilities, adherence to ethical principles, and sensitivity to a diverse patient population.
Systems-Based Practice	Residents must be able to demonstrate an awareness of and responsiveness to the larger context and system of health care and the ability to effectively call on system resources to provide care that is of optimal value.

Lurie demonstrates that, even if concepts like “systems-based practice” did not conflict with “professionalism”, none of the Core Competencies can be measured independently (32, 33). “Practice based learning” cannot be understood separately from “medical knowledge”, and “professionalism” cannot be interpreted outside the context of “patient care” or

“communication skills”, and so on. Because the CC fail to define separable and separating elements, none presents a valid concept. Instead, because they seem plausible, yet actually obscure what they intend to describe, the CC meet the definition of anti-concepts (34). Anti-concepts are not measurable. Attempting measurement certainly generates data, but those data are invalid, mapping to nothing real in the world (see below).

Even if one accepts the reductionist premise, the particular behaviors, or competencies, that comprise the Milestones do not sum to competence. Huddle and Heudebert point out that “anatomizing clinical competence according to discrete learning objectives does not and cannot add back together to constitute the skill and ability of the competent physician.” (35) Note the apt verb invoked here, suggesting an autopsy on a corpse: “anatomizing.” An illustration of this type of anatomizing is depicted in FIGURE 1. Because cataloguing disassembled parts is not a reliable method for interpreting the whole, any “attempt to capture higher-level clinical competence in objective measures is misconceived.” (35)

Here we come to a far deeper flaw in CBME: those “higher-level” cognitive capabilities, essential to mindful medical practice, are denied by the psychometric theory behind CBME. That theory is Behaviorism.

The language of the Milestones and Core Competencies, with relentless calls to “must be able to demonstrate,” reveals that CBME is explicitly Behaviorist, relying on premises and methods described by B.F. Skinner and others. Skinnerian Behaviorism reduces human minds to a black box, psychology to a kind of statistical process control, and human action to a series of “behaviors” driven solely by the “reinforcement” of external agents in the environment. Consider Skinner’s idea of education, in his own words: “A teacher arranges contingencies under which the student acquires behavior which will be useful to him under other contingencies later on.” (36) Human conscious agency is omitted by this theory of learning, not as irrelevant, but as non-existent. A recent review details and demonstrates the reliance of CBME on its Behaviorist roots (37).

While Skinner’s approach could describe some aspects of the innate behavior of animals, it offers scant insight into processes of human cognition (which it denied) or development of human proficiency (which it reduced to stimulus-response processes) (38). The Behaviorist lens of CBME is blind to critical elements of “the good doctor”, e.g. moral courage to act (open the chest, resist pressure to operate, etc); creative problem solving (in contrast with algorithm compliance); professional ownership (rather than shift work). CBME seems to envy statistical process control (39), the methods of which Skinner and others attempted to deploy, hoping to achieve for psychology the kind of rigor seen in physics. Behaviorism may have allowed some study of innate behaviors of rodents and pigeons, but no amount of physics envy allowed extrapolation to advanced human cognitive performance. Behaviorism could hardly be a less appropriate basis for teaching the very human, complex actions of successful surgical practice. Imagine teaching a resident to operate by arranging the proper “contingencies” that would shape her mindless responses.

Given this basis, it is less surprising that virtually no empirical support exists for CBME in the literature. Instead, nearly all publications on CBME resort to rhetorical advocacy or to appeals to authority: A recent review revealed that, of 2168 publications on the subject of competency- or outcomes-based medical education, just 8 were empirical studies, with weak or negative support for competency based methods. (37). The authors conclude that while CBME is vigorously advocated, this agenda "...has not been followed by substantial research on the impact of learning outcomes on teaching and learning medicine." CBME is not supported by science.

This lack of evidence does not deter advocates of CBME. "However, if formal evidence of the effectiveness of CBME is lacking...", it should be adopted anyway, states a typical propaganda piece (18). These authors claim that "sound advances in educational theory" underpin the "importance of clearly defined outcomes," but they never provide experimental support for competency-based methods. Those who demand the shift to CBME allow that there are serious "challenges" to implementing CBME, noting tangentially the huge time and money costs generated (see below) (24). Despite repeated promises of benefit, advocates never show actual payoffs, glossing over the costs while implying that they simply know CBME will deliver, eventually.

Perhaps CBME fails partly because it mistakes surgical training for a technical education. Of course, surgery depends on technical ability, the basic skills of stitching and cutting, the proper handling and use of particular tools and instruments, and the specific elements of particular procedures, etc. These all form the *techné* of surgery. *Techné* (Greek, "craft") gives us the root of technique. It is easy to be fooled into thinking that the lineaments of surgery are mainly "hand work" that can be measured with some kind of easily understood and "validated" scale. However, true surgical proficiency lies not in these aped behaviors, but in *metis*, the moral wisdom and insight gleaned from the contemplation of experience (40). *Metis* is found in the wry wisdom of the surgeon's old aphorism: "Good judgment comes from experience. Experience comes from bad judgment." In a surgical context, *metis* is the ability to judge, to integrate confusing and discordant clinical data, to know when not to operate, to have accurate self-knowledge of one's own technical ability for this patient, to sense trouble, to place countermeasures against misadventure, to rescue from unexpected problems, to have a backup plan...and another backup plan (41). CBME, in contrast, is all *techné* and no *metis*.

B. Bad Methods: A False Compass

Although a kind of Behaviorist statistical process control underlies it, the conceptual model of CBME renders measurement meaningless. Milestones collect numerical values and purport to measure educational outcomes through a kind of sampling (24). For instance, the Pediatric Surgery Milestones evaluations sample a handful of case types out of the hundreds of coded case types performed by pediatric surgeons. However, the case distribution of pediatric surgery, like other surgical specialties, is highly skewed, following a

classic long-tail distribution in which most case types are relatively rare ("We do a lot that we don't do much"). As a result, the product of the Milestones is not a statistical sampling of a specified aspect of a uniformly distributed and high-repetition process (such as, say, the tolerances in manufactured bolts), but a biased and inadequate sample. Even in a busy program with a small staff and a small number of trainees, the number of assessments of performance of, for example, "Wilm's tumor" in any given reporting period is miniscule, often none. The most common response becomes "unable to assess." This paltry sampling of such a broad field of human action cannot hope to measure much at all. Even with frequent evaluations, the data generated this way is far too sparse for legitimate sampling (42).

Worse, when the data are reported, the values must be aggregated and interpreted somehow. How? The ACGME does not say, except to demand that each program convene a "Clinical Competency Committee" that will figure it out in some kind of meeting. Unfortunately, the nature of the data produced resists analysis. First, an assumption of linearity in the scale is false. Like typical 5, 7, or 10 point scales which favor extreme answers (and therefore have a skewed distribution) (43), the Milestone scale is biased toward intermediate responses: the 0 and 5 scores are expected to be vanishingly rare (e.g. scores of 5 described as "aspirational.") Moreover, it is clear that the Milestones do not comprise interval scales—there is no expectation of regular steps between levels (44). How are we to aggregate and interpret this kind of data? Do we use means? Medians? ANOVA? Are any of these aggregation methods valid with these kinds of numbers? In fact, parametric statistical descriptors are not valid with this type of data (43). Mathematical methods fail (or must be abused) with sparse, non-monotonic, non-objective, non-parametric data like the Milestones yield.

Some argue that parametric statistical methods can still produce valid insights from nonparametric data (45), but there is another problem: Milestone data are often not even ordinal (44). For instance, some of the Pediatric Surgery Milestones (e.g. "Healthcare Delivery and Cost—Systems Based Practice"), reveal no logical progress among levels: "incorporates risk management in pediatric surgery" is Level 2 while "understands the principles of insurance coverage" is Level 3. Reversing the order of these "outcomes" would be just as plausible, rendering the notion of a student's progress here meaningless. Similarly, the Milestones for medical knowledge hold no hierarchy of knowledge, only a list of diseases grouped randomly. Other Milestone descriptions follow a similarly flawed construction.

This raises the problem of abstraction. Despite the essential idea that CBME depends upon specific, directly observable behaviors, Milestones descriptors rarely contain these. Faced with tension between concrete and abstract descriptions of behavior, the authors of Milestones hedge, wavering uncertainly among levels of abstraction. "Understands principles of..." is insufficiently concrete to be objective, while "demonstrates the ability to..." may be too narrow to be illuminating. Oscillating between airily abstract and unyieldingly granular, official Milestone descriptions already run for dozens of pages, but would need to go on for thousands of pages to enumerate all specifically observable

skills needed for all the variations of all the diseases presented to the physician (35). Instead, the burden of resolving the abstraction problem is shifted onto the evaluator, who defaults to a shrug and a guess.

This cost shift reveals the human-factors problem with CBME in general, and the Milestones in particular. Even if we stipulate that these scores could somehow map validly (if imperfectly) to surgical ability, the data still have a fatal reliability problem (46). Assessment forms are often regarded as tedious and confusing, leaving evaluators indifferent or even hostile to the process. The burden of mandated surveys and scores and evaluations leads quickly to fatigue. Fatigued evaluators answer hurriedly, even thoughtlessly (47). Harried evaluators shrug, roll their eyes, and draw a line straight through a series of questions, giving the same score on all assessments. These maladaptive and invalidating “response styles” are well-described (46), but are invited by the language and structure of the Milestones. Fatigue waylays evaluators who might otherwise attempt thoughtful responses, but even those who persist may be stayed by negative reactions to previous, more candid answers. It is common to hear complaints from those who have suffered personal penalties for providing honest but lackluster scores, and the Milestones replicate and magnify this experience (48).

The ACGME might argue that in its reported Milestone data, they see a broad spread on reported values for various trainees, suggesting that evaluators are not resorting to maladaptive response styles. However, tidily distributed scores do not support this conclusion. The Milestone data is not “raw”, but has been aggregated (somehow!) and reported by a small subset of evaluators, perhaps even by a single individual, at each institution. Whoever reports this data will reason that a sure way to imperil his training program would be to provide Milestone data that draws attention and official scrutiny, for example by being skewed or narrowed by indifferent responses. Better to “massage” that data a bit. In fact, a computer algorithm that randomly generated gently noisy data that corresponded roughly to trainees’ training levels would be the safest way to produce data for reporting. This manufactured data would have a reassuring distribution and reflect a general tendency to track with trainees’ progress. Authorities receiving this data would think all was well. The Milestones have an insurmountable garbage-in-garbage-out problem.

Analysis of this data can neither divide the able from the non-able, nor guide any individual toward competence. Indeed, the CBME-oriented Milestone scheme fails to meet any criteria for validity and reliability of psychometric measures (49). As a result, the process has already degenerated into little more than another form of non-value-added educational overhead, as both trainees and their teachers go through the motions in order to create a show of compliance for accreditors (48). These showpieces aside, nothing at all has been measured, and no trainee will have been guided from amateur to expert through a competency-based approach. Milestones therefore represent a false gradient from novice toward proficient. Objectives-based schemes frequently fail this way, and the more complex and creative the human task, the more likely that objective-setting will confuse, distract, distort, or subvert the product (50).

Strangely, while the ACGME mandates data reporting, the organization appears to have no plan for the data (24). There is no research program, no hypothesis, and no improvement theory. But data never speak for themselves; theory is always required to make sense of the data. The Milestone data, in contrast, are unsupported by a model. The implicit theory seems to be something like “measurement motivates mastery”, but the proponents of CBME have not suggested this or any other mechanism for how their predefined outcomes actually promote mastery. Instead, the real reason the data is demanded appears to be to force administrative *legibility* (40). Legibility is a form of homogenizing control over people or organizations by some central authority whose message is: “We are watching you.” Competencies are a cudgel, not a hornbook.

C. Bad Results: Costly Externalities

CBME carries serious and destructive costs, the kind of unintended consequences that economists euphemize as “negative externalities.” Most obvious among these costs are **expensive enforcement and compliance regimes**. While CBME promised to “reduce the administrative burden” (51) of accreditation, it has done the opposite. For example, at our own institution, administrative requirements for a tiny 3-trainee program now require a dedicated full-time coordinator, increasing non-clinical administrative time for the program director, more frequent demands for evaluations from the clinical staff, and increased contributions from many other non-clinical staff (e.g. business analysts, clerical staff, records workers, and research coordinators). An expensive subscription to customized residency management software is now required to catalogue, store, and track trainees’ data. Meanwhile, faculty face further non-clinical time burdens from requirements to attend meetings of new ACGME-mandated committees. Despite declarations that these methods would bring simplicity, savings and clarity (17), tangible and intangible administrative costs always increase. These costs would be paid gladly were the benefits apparent, but there is virtually no payoff for the trainees. With the candor one finds only in the very young or the very old, surgical trainees laugh at the Milestones with dismissive contempt. They don’t think they can learn surgery this way.

They cannot. Since implementation of the ACGME’s ideas, trainees have shown **marked diminishment in clinical readiness**. Many have remarked on the serious deterioration of professional attitude and personal ability manifested in surgical trainees since the ACGME’s policies have been implemented (52). Certainly, duty hour restrictions (DHR) have seriously damaged the training of surgeons (53). But DHR are only part of the problem. Program directors also recognize the irreconcilable task of both adhering to DHR and preserving clinical values imbedded in the Core Competencies (4). Meanwhile, surgeons everywhere worry about the palpable lack of readiness they now encounter in trainees: Medical students are no longer prepared to start residency, residents are no longer able to operate at an expected level of skill, and chief residents are no longer ready to begin practice or even begin a fellowship (52).

One might dismiss these worries as “days of giants” surgical nostalgia, or “kids these days” type grouching. But this argument is belied by widespread resort to remedial training such as boot camps, surgical simulation centers, focused courses like Fundamentals of Laparoscopic Surgery (FLS), and increased mentoring of junior faculty. These efforts do not reveal a profession laudably dedicated to constant improvement, but one desperately scrambling to bridge widening gaps in readiness. Naturally, these gaps are then cited by CBME proponents as justification for more adoption of competency-based methods to correct the inadequacies of surgical residency training (54). The toxin is offered as antidote.

Policies exert selection pressure. The 80-hour week, for example, selects for the least “professional”, penalizing personal ownership, rewarding the lazy, and placing the diligent into moral conflict (53). Similarly, a competency-oriented educational program selects for medical merit-badge seeking and indifferent box-checking by both student and teacher (33). Moreover, since competence consists of much more than competencies, trainees who have simply “checked off” the requirements are unlikely to actually be ready to perform well in the real world (35). Enforcement of CBME exerts pressure not toward proficiency, but toward incompetence.

It might be argued that CBME does not present any methods of teaching, only methods of assessment. However, CBME is explicitly imagined to guide the content of material in medical training. “Outcome-based education involves an educational approach in which the decisions about *the curriculum and evaluation are driven by the learning outcomes* that students should achieve.” (25)(emphasis added). Crafting a program to explicitly adhere to the competencies and other imperatives as advocates explicitly expect (55) is a form of the deplorable practice of teaching to the test. But the lurking threat of de-accreditation exerts enormous pressure to do just this. This threat brings up another externality, the **crushing of educational innovation** that accompanies the rigidly literal bureaucratic oversight required to implement a competency-based system.

Innovation, rather than being promoted as the ACGME claims to desire and reward (51), is instead extinguished. Program directors recognize the risk of crossing accreditors and imperiling their accreditation status. Occasionally, advocates of CBME strike a contrite tone regarding the inherent unworkability of CBME e.g. “multiple limitations have been recognized with the competencies (24).” Unfortunately, the enforcement apparatus of the ACGME does not evince the compensatory latitude such limitations demand. Instead, minor infractions attract official scrutiny, scrutiny which may bring expensive inquisition. As a result, innovation that could attract accreditors’ attention is seen as high-risk, but without any upside potential. By this anti-innovation effect, **minimal standards quickly become maximum performance**. All efforts are directed at appeasing (or fooling) (48) faceless, distant accreditors who can hide behind shields of bureaucratic privilege (e.g. closed meetings, unavailable meeting minutes, rules forbidding direct contact with residency review committee members, etc), and who are therefore essentially unaccountable. The dominant strategy

(56) for program coordinators and directors in this game becomes strict letter-of-the-law demonstrations of compliance.

Meanwhile, tension then arises between what practicing clinicians and better students sense about the nature of surgical expertise in the real world, and what regulators demand according to their constructs. As a result, bedside clinicians subvert the rules (as they do with work hour restrictions), or mutely deliver hollow shows of compliance. This is a quiet kind of work-to-rule strike, and is a prime mechanism by which utopian social systems engineering schemes fail (57). When competency-based behaviorist pedagogical philosophy is compelled by burdensome bureaucracy, educational progress is frozen, or even reversed.

To review, it has been shown that CBME is conceptually misguided, rooted in Behaviorism’s flawed model of human cognition and expertise. Attempting to measure outcomes using these concepts fails because the descriptors don’t map coherently to real things in the world. Even if the concepts were sound, the sampling is inadequate, and the data yielded are non-objective, non-monotonic, non-parametric, and even non-ordinal. Even if statistical interpretation was allowed here, the data are subverted at their source by human factors. The data therefore chart a false gradient toward proficiency, and create a false compass of educational progress. Finally, forcing programs to follow this false compass through administrative coercion damages real education, inflates educational costs, and destroys innovation.

But it doesn’t have to be this way.

IV. POSSIBLE ANTIDOTES

Traditional surgical education deserved to have its vices rooted out...but not its virtues. Humans do learn by imitation, immersion, and action, and this process is neither serial nor standardized. Human proficiency is not the crude sum of a series of islands, but an ever-more-elaborate self-reinforcing network of capabilities, the nodes of which can only be misunderstood in isolation. If the essential error in CBME is conceptual disintegration, it follows that training and evaluation that are conceptually integrated can offer solutions for both teaching and assessment.

Integrated teaching methods remain foreign to workaday “formal” instruction in medicine: The nested bullet-points of PowerPoint decks almost always lack themes, unifying ideas, guiding principles, and mental frameworks on which learners can cognitively hang details. Instead, reflecting the pervasive myths of OBE, speakers are compelled to produce and state some list of “learning objectives” for each lecture. These objectives go unread at best, or misunderstood at worst. Meanwhile, speakers proceed with a recitation of disconnected facts, “one damn thing after another,” a method famously pilloried by Edwin Tufte (58–60). Ironically, because humans are strongly inferential and imitative learners (61), and do what they have seen done, this fractured and ineffective teaching style persists. Learners taught with poor technique adopt these same poor techniques when they become teachers. Better methods exist.

A. Integrated Educational Methods

We can identify several mutually reinforcing elements in human learning outside of a “competency-based” paradigm. An integrated approach to surgical education would:

- Reveal the nature and use of practical principles that inform actions in a broad range of (but not all) circumstances. For example, the principle of “maximization of degrees of freedom” guides the learner toward a near-ideal trocar placement for any minimally invasive surgery case type. Application of this principle allows the learner to forego brute memorization of dozens of received arrangements for specific case types. The memorization strategy leaves the young surgeon disarmed before any novel circumstance; a principle-based strategy works generally (12). Meanwhile, the principle also reveals where a strategy will *not* work.

- Supply the “missing manual” for all the various types of equipment used in surgery. Everything from tubes to energy devices has an intended use designed into it, and when learners are ignorant (11) of these features, perhaps reasoning by analogy or best guess at how the thing works, they will misuse the device. This is the problem of “affordances” in design and use of devices (62), and presents a serious hidden threat to patients. For example, the Harmonic Scalpel™ (Ethicon; Sommerville, NJ) is often implicitly (but incorrectly) likened to monopolar cautery, and users who try to staunch bleeding from the surface of a liver or spleen with this device (the way they might with a monopolar energy tool) will find it ineffective at best, and catastrophic at worst. Explicit device instruction offers a powerful countermeasure to this type of hazard.

- Incorporate Ericsson’s ideas of deliberate practice (63). For some tasks that require dexterous manual facility, there is no substitute for coached shaping of motion, over many repetitions of supervised and unsupervised practice (63). This may be the real power of surgical simulation centers, which should be conceived more as batting cage than scrimmage field. Repeated drills and exercises using real instruments on models of task-specific fidelity (64) allow residents to come to the operating room with muscle memory for tying, needle handling, moving two hands, etc. This frees cognitive capacity for higher order thinking (and learning) in real cases (65).

- Update and expand working surgical traditions. For example, the semi-facetious ethic of “see one, do one, teach one” (66) was actually wrong in only one way. Specifically, it posits a far-too frugal experience with seeing, doing, and teaching. This is the joke of course: Practice is the most reliable means to achieve proficiency, and practice takes far more than one go (67). A better ethic would have been “**See some. Do many. Teach forever.**” But observe that See-Do-Teach does not place these activities in isolation. What do these terms mean, really? “See” means not idle watching, but mindful observing: What is the sequence? How exactly does the hand move? What settings? What set up? What do you do if this goes wrong? Etc. “Do” is not just aping these moves, but working for parsimonious action, assured tissue handling, ideal visualization. “Teach” integrates these experiences of

seeing and doing into the principles that inform a skillful operation. One teaches, ideally, not to see, but to identify what exactly to look for; not to do, but to find out exactly what methods and tricks and pitfalls make up the doing. All three actions recurse into each other.

- Explicitly direct training toward experiential learning that creates “gist” (68) or “recognition primed” (6) type of expertise. This human pattern-recognition super-power is probably what was intended by the inventors of “practice-based learning,” but they missed the mark. Instead, the essential means by which the truly proficient achieve this state is to “fill the experience bank” (6). That is, case exposure and Socratic questioning (another surgical tradition) over time build an ability to recognize complex patterns, to induce generalizations from few (or even one) examples (69), and to grasp nearly instantly what is really happening in a noisy environment. For example, experienced surgeons can often identify at a glance a patient in trouble, making both a correct diagnosis and clinical action plan in an instant. Learning from experience hones one’s “spidey sense” (70), the accurate (6) inner voice that says “trouble here!” But experiential learning does more than help keep surgeons out of trouble—it also gives them better ability to rescue patients from trouble, by giving them a mental “bag of tricks” to rely upon when misadventure visits. These abilities are the substance behind hackneyed buzzwords like *reliability* (reduce the probability of trouble) and *resilience* (increase the probability of recovering from trouble) that are mouthed by the same people who like to drone about “situational awareness.” Real experts don’t bother with these bureaucratic confections; they just see and do, proficiently.

- Incorporate “reflection in action” as described by Schön (71). Reflection in action might be described as professional practice with a recursive superego: the professional thinks about acting, and while acting, thinks about thinking about acting. She draws upon this reflection to monitor, adjust, and improve her performance, even as she is performing. Later, she thinks about this performance, remembering it during the next similar task, and scheming, quite specifically, how to improve (observe how instructors that use “the missing manual” type instruction can feed this process). This is not just careful thought, but the willful flow between thought and meta-thought, action and meta-action. This cognition-metacognition interplay is essential to the autodidact. Every proficient surgeon achieved proficiency through autodidactic cognitive progress using her experience as substrate (72).

- Accept the path-dependent nature of human learning. No two humans learn the same task the same way, and no two experts arrived at their proficiency by the same path. Two surgeons may both demonstrate stellar outcomes (short operating times, low infection rates, excellent cosmetic or functional results, et) but likely employ different techniques, tricks, and tools. These differences are acknowledged by the universal use of “preference cards” for individual surgeons. The notion of Milestones presupposes a single road, but that lone road is a myth. Every learner’s path is different. It follows that evaluation should be customized as well.

B. Integrated Evaluation Methods

Human performance in a given sphere of practice (like surgery) can and should be assessed according to methods that sum rather than divide. That is, evaluation methods should integrate the many networked threads that comprise proficiency, and report the results without resort to pseudonumerical scores. Some examples include:

- **Skill Evaluation:** The “Zwisch scale” is an integrated subjective assessment of surgical readiness. (73). The scale itself is ordinal, monotonic, intuitive, and validated. The evaluator, after direct observation of the resident in a case, rates them on their demonstrated ability to perform the operation with various levels of help from the attending surgeon: “show and tell,” “smart help,” “dumb help,” or “no help.” In addition to being inherently summative, the scale has the virtue of requiring no instruction, and evident face validity of the rankings averts confusion: Any practicing surgeon knows what those levels mean. Meanwhile, progression toward independence is easily demonstrated, and the scale can be adapted to surgical subfields (e.g. trauma) or even to individual case types. If combined with intermittent blinded rating of general technique (e.g. “tissue handling”, “parsimonious movement”, etc) (68), or even machine learning indicators of technical facility (74–76), faculty assessments of trainees’ technique would be simpler, more objective, and more actionable.

- **Narrative evaluations:** One of the unsubtle failures of the Behaviorist philosophy of evaluation is evaluator fatigue. Endless scales are dismissed as bothersome, irrelevant, or invalid by clinicians who must be nagged to complete the forms...then do so with indifference. In contrast, narrative evaluations that ask only one or two questions can be solicited infrequently. They don’t allow the evaluator to answer with a shrug and a check box. Instead, the answers are inherently explicit, descriptive, and impossible to be pseudonumerical. These kinds of answers allow training directors to hone in on specific problems that a trainee might have, and craft a custom plan for remediation. Accreditors disapprove of narratives because these answers are far less legible than number scales. Nevertheless, narratives can restore the rich contextual information content that lossy evaluation systems like Likert scales or Milestones strip away (77).

- **Knowledge Testing:** Of all the “core competencies”, the only one that has any claim to educational legitimacy is that of medical knowledge. Doctors need a vast mental store of internalized information, even in a Wikipedia age. More to the point, they need to be able to integrate that knowledge into contextual decisions (diagnosis) and actions (medical interventions) that will correctly move the patient from uncertainty (disease). Properly done, knowledge testing empirically correlates to actual performance, delineates a working body of knowledge required for a profession, and provides an objective standard for public accountability craved by regulators (78, 79).

The imperfections of these methods are their strength. While

integrative evaluation methods seem imprecise, they have the virtue of reflecting (rather than obfuscating) the real, multidimensional nature of human performance. Particularly when seeking to assess human action in clinical medicine, “the most complex task humans ever attempt” (80), the oversimplifications of competency-based methods sacrifice educational insight on an altar of administrative legibility. In other words, regulators act as if they would rather have their tallies and scales because these look like data and are easier to count—and to control—than the messy idea of the able surgeon. But look again at Figure 1. Summing up is superior to breaking down.

V. CONCLUSIONS

Milton Friedman once said on television: “One of the great mistakes is to judge policies and programs by their *intentions* rather than their *results*” (emphasis added). Advocates for CBME mistake intention for results. They intended objective measurement of medical training and, through this, improvement of doctors in general; the results include confusion, disarray, massive administrative waste, disdain for authority (81), and educational decay. CBME promised to create “objective outcomes” and “standards” in medical training that improve “accountability” for “public stakeholders”. Instead, following a competency-based philosophy, the ACGME has been experimenting on a highly-complex system using unproven, nontrivial interventions on unconsented subjects. This educational Lysenkoism (82) endangers patients, raises costs and bureaucratic burdens, and leads to deskilling in the trainees. Left unchecked, these methods will upend surgical education in a single generation.

CBME has displayed the “behavior” of the classic hype cycle (83) at its “peak of inflated expectations”: Widespread and enthusiastic calls for embracing the visionary new paradigm, appeals to authority substituted for empirical demonstration, amplified references in media, dedicated conferences, reams of descriptors and requirements, plenty of excitement...but virtually no substance. The ascent of CBME has been aided by a glamour of reasonableness, what Schön identified as the epistemic trap of “Technical Rationality” (71). But CBME is scientism, not science, and it doesn’t work; humans do not learn according to the cartoonish behaviorism underlying CBME. Surgeons certainly cannot practice this way (68).

The hype could burble back down to a productive simmer as people discarded the unworkable and the fantastic, and learned to profitably exploit the narrow applications for which objectives-based education could be useful. But with the enforcement apparatus of ACGME accreditation cementing this impoverished pedagogy in place, CBME will continue to damage and degrade surgical education. A surgeon is not a pediatrician is not a cardiologist—superficial similarities imagined in the core competencies and overwrought common program requirements mislead more than guide. Forcing this homogeneity by means of the CBME sieve drives the rich variety of surgical expertise into an uncanny valley of lowest-common-denominators. CBME is the great orbitoclast of surgical training.

Surgical organizations like the American College of Surgeons (ACS) and the American Board of Surgery (ABS) must rescue their roles as leaders in surgical education. The ACGME holds monopoly power in surgical education because both the ABS and ACS have ceded control and theoretical exposition in surgeons' education. Like all monopolies, this one damages the field it occupies. And like all monopolies, this one is artificially maintained in part by coercion (e.g. the ABS rules that require candidates for their examinations to have attended an ACGME accredited program), and in part by the tacit sanction of its victims. In other words, CBME is unsalvageable, but is sustained in an undead form by the (unwitting) contributions of surgeons' personal efforts. This energy, rather than guiding trainees, instead has been coopted, and now only protects the magical thinking of the competency-based paradigm.

Unequivocal repudiation of the competency-based dogma might seem extreme, but such a rejection would instead be a responsible exercise in professional accountability. In any case, official pushback would be preferable to limply accepting the imbecility of "entrustable professional activities" or the futility of surgical training through bureaucratic behaviorism. CBME is wrong for surgery. Surgeons must choose: Protect, preserve, and build the profession, or accede to its destruction through mute acceptance of an unworkable educational fad.

Fig. 1. Conceptual disintegration is illustrated in Ursus Wehrli's reinterpretation of a branch. Biologically, the integrating principle of the branch ("Before") is maximization of surface area using an energy minimizing algorithm that is a form of biological fractal. Disintegration seemingly renders the branch easily legible ("After") but at the cost of massive information loss. Certainly elements (number of spines and twigs) can be counted, yet these elements do not count.

REFERENCES

1. Cameron, J. L. 1997. William Stewart Halsted. Our surgical heritage. *Annals of surgery*
2. Mayberry, J. C. 2003. Residency reform Halsted-style. *Journal of the American College of Surgeons*
3. Ahmed, N., K. S. Devitt, I. Keshet, J. Spicer, K. Imrie, L. Feldman, J. Cools-Lartigue, A. Kayssi, N. Lipsman, and M. Elmi. 2014. A systematic review of the effects of resident duty hour restrictions in surgery: impact on resident wellness, training, and patient outcomes. *Annals of surgery* 259: 1041-1053.
4. Antiel, R. M., S. M. Thompson, F. W. Hafferty, and K. M. James.... 2011. Duty hour recommendations and implications for meeting the ACGME core competencies: views of residency directors. *Mayo Clinic* ...
5. Lewis, F. R., and M. E. Klingensmith. 2012. Issues in general surgery residency training--2012. *Ann Surg* 256: 553-559.
6. Klein, G. 1999. Sources of power: How people make decisions. MIT press,
7. Ericsson, K. A. 2008. Deliberate practice and acquisition of expert performance: a general overview. *Academic emergency medicine: official journal of the* ...
8. Hugh, T. B. 2002. Laparoscopic bile duct injury: some myths. *ANZ journal of surgery*
9. Lichtenstein, I. L., and J. M. Shore. 1976. Exploding the myths of hernia repair. *The American Journal of Surgery*
10. Blinman, T., and M. Maggard. 2000. Rational manipulation of oxygen delivery. *J Surg Res* 92: 120-141.
11. Watanabe, Y., Y. Kurashima, A. Madani, L. S. Feldman, M. Ishida, A. Oshita, T. Naitoh, K. Noma, K. Yasumasa, H. Nagata, F. Nakamura, K. Ono, Y. Suzuki, N. Matsuhashi, T. Shichinohe, and S. Hirano. 2016. Surgeons have knowledge gaps in the safe use of energy devices: a multicenter cross-sectional study. *Surg Endosc* 30: 588-592.
12. Blinman, T. A. 2015. MIS-behavior: practical heuristics for precise pediatric minimally invasive surgery. *Urol Clin North Am* 42: 131-140.
- 13...., M. W. Feldman, L. Fogarty, S. Ghirlanda, and T. Lillicrap.... 2010. Why copy others? Insights from the social learning strategies tournament. ...
- 14.Jr, W. P. R. 2004. Basic certification in surgery by the American Board of Surgery (ABS): what does it mean? Does it have value? Is it relevant? A personal opinion. *Annals of surgery*
- 15.Orgill, B. D., and D. Simpson. 2014. Toward a glossary of competency-based medical education terms. *Journal of graduate medical education*
- 16.Crebbin, W., and D. Watters. 2012. Defining Standards for a Competency-Based Surgical Training Program. *Journal of graduate medical education*
- 17.Byrne, L. M., R. S. Miller, and T. J. Nasca. 2016. Implementing the Next Accreditation System: Results of the 2014-2015 Annual Data Review. *J Grad Med Educ* 8: 118-123.
- 18.Carraccio, C., R. Englander, E. Van Melle, O. ten Cate, J. Lockyer, M.-K. Chan, J. R. Frank, and L. S. Snell. 2016. Advancing Competency-Based Medical Education: A Charter for Clinicianâ "Educators. *Academic Medicine* 91: 645-649.
- 19.Harden, R. M. 2007. Outcome-based education: the future is today. *Medical teacher* 29: 625-629.
- 20.Harden J R Crosby M H Davis M Friedman, R. M. 1999. AMEE Guide No. 14: Outcome-based education: Part 5-From competency to meta-competency: a model for the specification of learning outcomes. *Med Teach* 21: 546-552.
- 21.Frank, J. R., R. Mungroo, Y. Ahmad, M. Wang, S. De Rossi, and T. Horsley. 2010. Toward a definition of competency-based education in medicine: a systematic review of published definitions. *Med Teach* 32: 631-637.
- 22.Cate, O. T., and F. Scheele. 2007. Viewpoint: competency-based postgraduate training: can we bridge the gap between theory and clinical practice. *Academic Medicine*
- 23.Batalden, P. B., and D. C. Leach. 2009. Sharpening the focus on systems-based practice. *J Grad Med Educ* 1: 1-3.
- 24.Potts, J. R. 2016. Assessment of Competence: The Accreditation Council for Graduate Medical Education/Residency Review Committee Perspective. *Surgical Clinics of North America*
- 25.Shumway, J. M., R. M. Harden, and F. M. E. I. E. Association. 2003. AMEE Guide No. 25: The assessment of learning outcomes for the competent and reflective physician. *Med Teach* 25: 569-584.
- 26.Harden, R. M. 2014. Progression in competency-based education. *Med Educ* 48: 838.
- 27.Ben-David, M. F. 1999. AMEE Guide No. 14: Outcome-based education: Part 3-Assessment in outcome-based education. *Medical Teacher* 21: 23-25.
- 28.Peikoff, L. 2013. The Dim Hypothesis: Why the Lights of the West Are Going Out. NAL,
- 29.Wasnick, J. D., L. Chang, and C. Russell.... 2010. Do residency applicants know what the ACGME core competencies are? One program's experience. *Academic Medicine*
- 30.Bernstein, J. 2014. Not the last word: The ACGME core competencies are overrated. *Clinical orthopaedics and related research* 472: 2932.
- 31.Hafferty, F. W., and R. Franks. 1994. The hidden curriculum, ethics teaching, and the structure of medical education. *Academic Medicine* 69: 861-871.
- 32.Lurie, S. J., C. J. Mooney, and J. M. Lyness. 2009. Measurement of the general competencies of the accreditation council for graduate medical education: a systematic review. *Academic Medicine* 84: 301-309.

33. Lurie, S. J., C. J. Mooney, and J. M. Lyness. 2011. Commentary: pitfalls in assessment of competency-based educational objectives. *Academic Medicine* 86: 412-414.
34. Rand, A. 1990. *Introduction to Objectivist Epistemology: Expanded Second Edition*. NAL.
35. Huddle, T. S., and G. R. Heudebert. 2007. Viewpoint: Taking apart the art: The risk of anatomizing clinical competence. *Academic Medicine* 82: 536-541.
36. Skinner, B. F. 2011. *About behaviorism*. books.google.com
37. Morcke, A. M., T. Dornan, and B. Eika. 2013. Outcome (competency) based education: an exploration of its origins, theoretical basis, and empirical evidence. *Adv Health Sci Educ Theory Pract* 18: 851-863.
38. Seligman, M. E., P. Railton, R. F. Baumeister, and C. Sripada. 2013. Navigating Into the Future or Driven by the Past. *Perspect Psychol Sci* 8: 119-141.
39. Shewhart, W. A., and W. E. Deming. 1939. *Statistical method from the viewpoint of quality control*. books.google.com
40. Scott, J. C. 1999. *Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press.
41. Flin, R., G. Youngson, and S. Yule. 2007. How do surgeons make intraoperative decisions. *Quality and Safety in Health* ...
42. Ercan, I., Y. Yang, G. Özkaya, S. Cangur, B. Ediz, and I. Kan. 2008. Misusage of statistics in medical research. *tSPACE.library.utoronto.ca*
43. Jamieson, S. 2004. Likert scales: how to (ab) use them. *Medical education*
44. Stevens, S. S. 1946. On the theory of scales of measurement. *psychology.okstate.edu*
45. Norman, G. 2010. Likert scales, levels of measurement and the "laws" of statistics. *Adv Health Sci Educ Theory Pract* 15: 625-632.
46. Vaerenbergh, Y. V., and T. D. Thomas. 2013. Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion* ...
47. Eggleston, B. L., S. M. Miller, and N. J. Meropol. 2011. The impact of misclassification due to survey response fatigue on estimation and identifiability of treatment effects. *Statistics in medicine* 30: 3560-3572.
48. Williams, R. G., G. L. Dunnington, J. D. Mellinger, and D. L. Klamen. 2015. Placing constraints on the use of the ACGME milestones: A commentary on the limitations of global performance ratings. *Academic Medicine* 90: 404-407.
49. Cook, D. A., and T. J. Beckman. 2006. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med* 119: 166.e7-16.
50. Stanley, K. O., and J. Lehman. 2015. *Why Greatness Cannot Be Planned: The Myth of the Objective*. Springer.
51. Nasca, T. J., I. Philibert, T. Brigham, and T. C. Flynn. 2012. The next GME accreditation system—rationale and benefits. *New England Journal of Medicine* 366: 1051-1056.
52. Mattar, S. G., A. A. Alseidi, D. B. Jones, D. R. Jeyarajah, L. L. Swanstrom, R. W. Aye, S. D. Wexner, J. M. Martinez, S. B. Ross, and M. M. Awad. 2013. General surgery residency inadequately prepares trainees for fellowship: results of a survey of fellowship program directors. *Annals of surgery* 258: 440-449.
53. Antiel, R. M., and T. A. Blinman. 2016. To Leave or to Lie: Duty Hour Restrictions and Patient Ownership. *The American Journal of Bioethics*
54. Evans, C. H., and K. D. Schenarts. 2016. Evolving educational techniques in surgical training. *Surgical Clinics of North America*
55. Harden, R. M. 2007. Outcome-based education—“the ostrich, the peacock and the beaver”. *Medical Teacher* 29: 666-671.
56. Luce, R. D., and H. Raiffa. 2012. *Games and decisions: Introduction and critical survey*.
57. Bar-Yam, Y. 2003. When systems engineering fails—toward complex systems engineering. *Systems*
58. Tufte, E. R. 1997. *Visual explanation. Images and Quantities*
59. Tufte, E. R. 2003. *The cognitive style of PowerPoint*. Graphics Press Cheshire, CT
60. Tufte, E. R. 2006. *Beautiful evidence*. New York
61. Tenenbaum, J. B., C. Kemp, and T. L. Griffiths. 2011. How to grow a mind: Statistics, structure, and abstraction. *science*
62. Norman, D. A. 2013. *The design of everyday things: Revised and expanded edition*. Basic books.
63. Ericsson, K. A., N. Charness, P. J. Feltovich, and R. R. Hoffman. 2006. *The Cambridge handbook of expertise and expert performance*. Cambridge University Press.
64. Edmunds, T., and D. K. Pai. 2012. Perceptually Augmented Simulator Design. *IEEE Trans Haptics* 5: 66-76.
65. Economides, M., Z. Kurth-Nelson, A. Lübbert, M. Guitart-Masip, and R. J. Dolan. 2015. Model-Based Reasoning in Humans Becomes Automatic with Training. *PLoS Comput Biol* 11: e1004463.
66. Kotsis, S. V., and K. C. Chung. 2013. Application of See One, Do One, Teach One Concept in Surgical Training. *Plastic and reconstructive surgery*
67. Colvin, G. 2008. Talent is overrated: what really separated world-class performers from everybody else.
68. Reyna, V. F. 2008. A theory of medical decision making and health: fuzzy trace theory. *Medical Decision Making*
69. Lake, B. M., R. Salakhutdinov, and J. B. Tenenbaum. 2015. Human-level concept learning through probabilistic program induction. *Science*
70. Bond, S., and S. Cooper. 2006. Modelling emergency decisions: recognition-primed decision making. The literature in relation to an ophthalmic critical incident. *Journal of clinical nursing*
71. Schon, D. 1983. *The reflective practitioner*. New York: Basic Books
72. Harrison, M. R., and N. S. Adzick. 1990. Advice to a resident: a surgeon's notebook. *J Pediatr Surg* 25: 379-380.
73. DaRosa, D. A., J. B. Zwischenberger, and S. L. Meyerson. 2013. A theory-based model for teaching and assessing residents in the operating room. *Journal of surgical* ...
74. Megali, G., S. Sinigaglia, and O. Tonet. 2006. Modelling and evaluation of surgical performance using hidden Markov models. *IEEE Transactions on* ...
75. Rosen, J., B. Hannaford, and C. G. Richards. 2001. Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills. *IEEE transactions on* ...
76. Rosen, J., J. D. Brown, and L. Chang. 2006. Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete Markov model. *IEEE Transactions* ...
77. Ginsburg, S., J. McIlroy, O. Oulanova, K. Eva, and G. Regehr. 2010. Toward authentic clinical evaluation: pitfalls in the pursuit of competency. *Acad Med* 85: 780-786.
78. Haladyna, T. M., and G. A. Kramer. 2004. The validity of subscores for a credentialing test. *Eval Health Prof* 27: 349-368.
79. Haladyna, T. M. 2012. Developing and validating multiple-choice test items.
80. Kizer, K. 2000. Ten steps you can take to improve patient safety in your facility. *Brief Patient Saf*
81. Keizer, K., S. Lindenberg, and L. Steg. 2008. The spreading of disorder. *Science*
82. Hagemann, R. 2002. How did East German genetics avoid Lysenkoism. *TRENDS in Genetics*
83. Fenn, J. 1995. When to leap on the hype cycle. *Gartner ID: SPA-ATA-305*