Human Factors Tools for Improving Simulation Activities in Continuing Medical Education

F. Jacob Seagull, PhD

Human factors (HF) is a discipline often drawn upon when there is a need to train people to perform complex, high-stakes tasks and effectively assess their performance. Complex tasks often present unique challenges for training and assessment. HF has developed specialized techniques that have been effective in overcoming several of these challenges in work settings such as aviation, process control, and the military. Many HF techniques could be applied to simulation in continuing medical education to enhance effectiveness of simulation and training, yet these techniques are not widely known by medical educators. Three HF techniques are described that could benefit health care simulation in areas of training techniques, assessment, and task design: (1) bandwidth feedback techniques for designing better feedback and task guidance, (2) dual-task assessment techniques that can differentiate levels of expertise in tasks where performance is essentially perfect, and (3) task abstraction techniques for developing task-relevant fidelity for simulations. Examples of each technique are given from work settings in which these principles have been applied successfully. Application of these principles to medical simulation and medical education is discussed. Adapting these techniques to health care could improve training in medical education.

Key Words: *simulation, human factors, feedback, task analysis, dual-task paradigm*

Introduction

Human factors (HF) is the scientific discipline concerned with how people interact with each other and with technology in work settings. It is a profession that applies scientific theory, principles, data, and other methods to optimize human performance within those settings.[1] HF professionals have been working for over a century on measuring human performance and improving human ability in the workplace. As a field, HF is focused on issues of human performance in complex environments, striving to improve the ability of people to perform safely and efficiently through the design of technology, tasks, the environment, and training techniques.

HF is often applied to the issue of effective training. One of the early, simple examples of HF improving training is the work of Gilbreth, one of the fathers of HF. In the early 20th century, he devised a standard set of mnemonics (memory devices) for teaching Morse code to military personnel, intended to reduce the time needed to learn the codes.[2] Instead of memorizing the seemingly arbitrary combinations of dots and dashes representing the letters, Gilbreth created a list of words whose accents corresponded to the code for the various letters. "A" was "a-BOUT" (dot-DASH), "B" for "BOIS-ter-ous-ly" (DASH-dot-dot-dot), C was "CARE-less CHILD-ren" (DASH-dot-DASH-dot), D was DAN-ger-ous (DASH-dot-dot), and so on. The technique of generating mnemonics for aiding memory may be familiar to doctors who remember the 5 components of the Apgar score (named after Dr. Virginia Apgar[3]), by the 5 letters in her name: **A**ppearance (skin color), **P**ulse (heart rate), **G**rimace (reflex irritability), **A**ctivity (muscle tone), and **R**espiration.

HF was also heavily involved with the development, validation, and use of aviation simulators.[4] The extensive work in the area of simulation yielded lessons regarding the effective use of simulators, as well as techniques for assessing performance and improving learning. Although these techniques are widely used by HF professionals, they have not been widely adopted in the medical education community. HF is not well known or well integrated into medical education. However, both fields focus on high-stakes environments and the acquisition of complex skills.

This article describes 3 specific techniques that have each been applied successfully in other domains and would be

applicable to physicians at all stages of their education, including continuing medical education (CME). They were chosen because they are well known and well used in HF but virtually unknown in medical education. The techniques have the potential for addressing perennial problems in medical education for the design, measurement, and instructional techniques associated with complex medical manual and cognitive tasks. The 3 techniques are:

- *Bandwidth feedback,* a technique for presenting learners with feedback during performance only when appropriate in order to foster effective performance strategies. It has been found to improve the transfer of skills from a training environment to real-world performance.
- *Dual-task paradigm,* which uses an indirect means to assess performance on tasks where direct measures are not effective. It can provide more accurate assessment of expertise in complex skills.
- *Task abstraction* concepts, which can provide insight into relevant aspects of fidelity and validity in training paradigms.

These techniques would be particularly useful to people concerned with developing methods to teach complex manual skills and assess performance on these skills.

Three Techniques

Bandwidth Feedback

Skill acquisition often requires guidance during the initial stages. Successive approximation techniques may be used in which teachers may initially provide learners simplified tasks or subsets of the full task to master.[5] When some level of proficiency is gained, more complex tasks are provided. An example of this training strategy is used when learning how to ride a bicycle: a tricycle is first ridden, then a bicycle with training wheels, then the training wheels are removed.

A goal during the development of expertise is for the learners to learn to extract from the environment the information that is most relevant to the task. They learn to process that information and to derive the appropriate course of action based on the information. Thus, one aspect of skill acquisition is learning to distinguish the relevant cues in the environment to guide action. During the initial stages of skill acquisition, guidance, often in the form of positive and/or negative feedback, is provided to foster the acquisition of the requisite skills. However, feedback can also have adverse effects.

Aviation provides a human factors example of these adverse effects. In flight training, simulators have been used for teaching pilots to maintain proper speed, heading, and approach angle for final approach. To help speed learning, the simulators provided auxiliary information and feedback that were not provided in actual flight. Synthetic displays of the optimal flight-path for landing could be displayed

continuously on a flight simulator to help guide the learner on the optimal flight path. The pilot could see the through-the-window view of the approach, and additionally, avionic "symbology" (the collection of iconic symbols representing flight path, horizon, attitude, etc.) could be superimposed on the scene to show the pilot's deviation from the prescribed flight path. In contrast to the simulator, nonsimulated (real) flight contained no such displays. Pilots trained with the artificial guidance learned to fly simulators more quickly than those training without such guidance.[6] However, when transferring from the simulator to the operational flight deck, they could not perform as well as those trained without such guidance. Pilots trained using this artificial display became dependent on it and learned to ignore relevant cues from the environment. The feedback facilitated initial learning, but skills did not transfer well to the nonsimulated task.[7]

To resolve this conundrum of transfer of training, a new "bandwidth" technique was formulated for providing real-time feedback. The term *bandwidth* is used to describe a range ("band") of performance, or a tolerance for error. Feedback is given when performance falls outside of the normal bandwidth of performance. For example, a "band" of road has rumble strips at its outside margins. As long as the driver remains within the bandwidth of the highway, no feedback is given. The technique of bandwidth feedback is well suited for continual navigation or control tasks, where there is a continuous "correct" position or state, such as flexible endoscopy, bronchoscopy, or in the monitoring of a process. For flight training, when pilots flew close to the targeted flight path, no display augmentation was presented. Only when the pilot deviated a predetermined distance from the proper path would the avionic symbology appear to help correct the pilot's performance. When performance returned to acceptable levels, the symbology would disappear.

Thus, the pilots learned to rely on the cues in the environment, and not the display augmentation, to develop their skills. Only when they were unable to perform unassisted was augmentation presented. As the learner moved from novice to proficiency, the supplemental feedback would appear less and less, until the pilot could fly entirely unassisted.

With a growing emphasis on simulation in teaching medical skills, the human factors lessons from aviation simulation may prove useful. Current practices in surgical simulation provide examples of productive and suboptimal use of feedback, based on the principle of bandwidth feedback. Two examples of exercises in virtual reality training for laparoscopic simulators demonstrate the differences in feedback strategy. There is a skill-building exercise in the LapMentor simulator (Symbionix Co, Cleveland, OH) in which the trainee must stretch a vessel a prescribed amount. Prospective task guidance is given (akin to continuous, on-task feedback), and the feedback (guidance) is necessary to complete the task. The vessel changes to a green color to indicate that

it has been stretched the appropriate amount. Other exercises include markers such as green hash marks superimposed on the represented anatomical structures, indicating where they should be grasped, cut, or stapled. Thus, the users are taught to pay attention to the symbols and markers, not the anatomical structures. An alternative approach akin to bandwidth feedback applied to these examples would provide feedback only when the trainee surgeon lacerates the liver or otherwise injures tissue. Providing feedback only when (or slightly before) injury occurs does adhere to the principle of providing feedback only when the learner deviates from the optimal, in order to shape behavior away from negative outcomes such as liver injury.

Bandwidth feedback is not preferable in all situations. Continuous task guidance can be entirely appropriate for training novices. As skill progresses, bandwidth feedback techniques can be used to wean learners from task guides while still facilitating good performance strategies through feedback when poor strategies or performance occurs. More generally, the principle of providing diminishing levels of feedback as the learner advances has been shown to be effective even without tying the onset of feedback to how well the learner is doing.[8] It is an effective principle and widely accepted in medical education. Yet bandwidth feedback is qualitatively different than simply providing less feedback to trainees as learners advance, and bandwidth feedback has been shown to be a more effective technique[8,9] for facilitating transfer of training to the targeted task, and superior performance on the target skill.

Dual-Task Paradigm

Medical education is often faced with the challenge of measuring the performance of experts on complex tasks. Measurement can be difficult, as the criteria for expert performance are sometimes not defined operationally. Speed and accuracy measures are most often used to assess task performance, and in many cases, provide a good balance between measurement ease and the measure's discrimination ability. However, there often arise situations in which these types of measures are not sufficient:

- In certain tasks where time measures are not meaningful, such as when monitoring a patient. There is no metric for "speed" of monitoring, except when detecting changes or reporting results of monitoring.
- Where quantitative metrics of task performance are not available (such as surgical tasks for which measures of skill are not agreed upon or where outcome measures may confound surgical skill with variability in patient acuity).
- Where metrics do not sufficiently differentiate between skill levels—when score of an advanced beginner on a task is no different from that of an expert.

In addition to standard, objective performance metrics for a given task, there are techniques for using valid and reliable subjective measures, as well as a number of possible technology-intensive observation techniques (such as video-based motion analysis or eye tracking), and physiological measures (such as heart rate, brainwave patterns, catecholamine levels) to assess the difficulty an operator has in completing a given task. These methods also have shortcomings. For example, subjective measures are subject to biases,[10] and physiological measures are often cumbersome and resource-intensive.[11]

To counter the shortcomings of the above-mentioned techniques, HF researchers have adopted a technique pioneered in cognitive psychology for assessing performance indirectly, through assessment of workload. The concept of workload centers on the relationship between the individual's abilities and the demands of the task being performed. Workload, or the degree to which a person's capacity to perform a task is being taxed, can be assessed most simply by asking the person performing the task to judge how difficult the task was. There are more sophisticated, validated measures of workload that rely on subjective assessments on task dimensions such as mental, physical, temporal demands, as well as performance, effort, and frustration levels.[12] Beyond subjective measures, there is a paradigm for measuring workload through objective performance measures, known as a dual-task paradigm, or concurrent task paradigm.

Performance metrics on a given single task of interest may be at or near ceiling levels before the learner is fully competent. Performance metrics may not reflect a difference between novice learners who manage to succeed with difficulty, and those with greater proficiency who succeed with ease. The dual-task paradigm can be used in these situations to provide a proxy measure of expertise, and differentiate between struggling novice and coasting expert who each achieve the same observable performance on the task of interest.

The demands for a single task may not be sufficient to impair performance of the primary task of interest. However, the ability to complete a second, additional task concurrently can be a very sensitive measure of the level of expertise. Although each task may be completed with 100% accuracy when performed individually, performing both tasks concurrently results in less than 100% accuracy. The composite score derived from both tasks concurrently is often sensitive to differences between different levels of expertise.[13]

This dual-task technique was developed to investigate the structure of human mind, explore the nature of human attention, and determine which tasks can be accomplished simultaneously. When performed simultaneously, tasks can compete with one another or can be accomplished with little interference. The cliché about walking and chewing gum provides an example of two tasks that can be accomplished simultaneously by all but those with minimal ability.

In contrast, the ability to transcription-type is disrupted by conversation when the typist is inexperienced, but not when the typist is experienced.[14,15]

Simple tasks such as toe tapping,[16] prospective time estimation,[17] or mental math are all techniques that can be added to many types of task in order to estimate expertise when other measures of performance may not successfully differentiate between novice and expert. These techniques have been applied occasionally within the domain of medical education and medical human factors research. Weinger[18] studied the workload experienced by anesthesia care providers, and found that a provider's ability to simultaneously provide anesthesia to a patient and detect the illumination of a red indicator light placed in the anesthesia workspace was impaired when the provider was under stress or high workload. More recently, Stefanides et al[19] applied the dual-task technique to skill acquisition in laparoscopic surgery tasks, showing that dual-task scores could discriminate between proficient and expert performance more effectively than single-task measures.

While this dual-task technique has been applied successfully in some areas of medical education, its use is still limited. The dual-task should be considered a viable technique for assessment of tasks in which performance measures are not readily available. If measures are not sensitive to expertise levels, consider adding a simple secondary task and measuring the combined score or performance level.[20] Criteria for secondary-task selection are beyond the scope of this article, and selection of appropriate tasks may require some effort. However, the technique can be valuable and provide more accurate metrics for assessment of expertise.

Task Analysis and Task Abstraction

A common approach to training medical tasks is to attempt to replicate the operational environment as closely as possible. The concept of "fidelity" in training environments is often confounded with face validity—the subjective estimation of whether a task or simulator or training method "appears" to be valid. Face validity has been often used as a criterion for assessing simulators used in medical education. Often, surface characteristics, such as visual fidelity in virtual reality (VR) simulators, are a strong influence on the selection of pedagogical techniques. Indeed, fidelity is important; however, visual fidelity and other superficial aspects of tasks comprise only one aspect of task fidelity.[21]

More important than attaining a close match between the appearance of the training environment and the operational environment is understanding the relevant aspects of the task that require fidelity. Educators may be tempted to choose methods to train target skills by evaluating the surface similarities between the targeted task and the training analog of the task. However, understanding the underlying structure of the tasks involved is essential.[22] Simulation can be understood on physical, semantic, and phenomenological levels.[23] *Physical* fidelity is associated with the physical properties such as size, weight, texture, and motion of the simulation. *Semantic* fidelity relates to the concepts and meaning associated with the simulation—physiological concepts such as hemodynamics, or the concept of a broken bone, independent from the modality that the concept is represented in the simulation (eg, through a mannequin, anatomical model, verbal or written description). *Phenomenological* fidelity concerns the way that the simulation is experienced by the individual in terms of emotion, thoughts, and beliefs—such as whether learners felt the simulation was realistic, valuable, and effective.

There is a set of techniques that facilitate an understanding of tasks at these and other deeper levels.[22,23] Human factors provides taxonomies of physical and semantic task structure (motor, visual, spatial, verbal, lexical, etc).[23,24] There are means of analyzing the semantic structure of knowledge within a domain for a given task (means-ends hierarchy,[25,26] work domain analysis.[27]) The underlying skill or strategy can be developed through skills that do not appear on the surface to be related to the targeted task. Thus, when discussing the realism or "fidelity" of a simulation, it becomes important to understand that the 3 levels of fidelity (physical, semantic, and phenomenological) can be largely independent of one another. Physical fidelity may not be related to the phenomenological fidelity, and phenomenological fidelity may not be related to the effectiveness of a simulator.

Aviation addressed questions of simulators' physical realism in the use of motion platforms in flight simulators.[28] The idea of platform motion appealed to the sense of physical and phenomenological validity, but validation studies showed that platform motion often did not improve transfer to operational environments. Pilots could learn to associate certain maneuvers with particular simulator motion, which did not adequately represent the g-forces involved in true flight, leading to performance that was poorer than pilots trained in stationary simulators—training with no motion was more effective than training with imperfect motion. Thus, simulators that deviated from "realism" in appropriate ways could be more effective than more "realistic" ones.[28] Similar questions of realism and fidelity are arising with respect to haptic (touch) feedback in VR surgical simulators.[29] Haptic feedback is imperfect (but evolving and improving), and it is not clear whether imperfect haptic feedback is more beneficial for learning than no haptic feedback.

There are examples of successful simulations for training that have neglected the physical fidelity, and instead focused on the semantic and phenomenological fidelity through semantic abstraction of the training task. An informative example of "task abstraction" in the domain of flight simulation can be seen in a Defense Advanced Research Projects

Agency (DARPA) project called "Space Fortress."[30] The task of flying a high-performance military jet was analyzed by human factors psychologists and cognitive psychologists in terms of basic perceptual and cognitive requirements. The abstract requirements that were generated were then used as a set of guidelines to design a video game similar to Space Invaders™ or Asteroids™ with no direct tie to flying a jet. This simple video game had limited fidelity when judged by face validity. The game interface consisted of simple, low-resolution graphics. A stick-figure spaceship was "flown" around a computer screen, circling a "space fortress." The player was required to avoid and destroy mines, and avoid getting shot by the enemy ship, while attempting to destroy the space fortress. It appeared to be a simple video game. However, it was also an excellent trainer for attention, memory, and workload management. To succeed in this video game, the player needed to demonstrate eye-hand coordination and manual dexterity for flight control and aiming weapons. Distinguishing friends from foes and activating weapons systems required both long- and short-term memory.

Despite the limited face validity of the video game, it was a very effective trainer in military flight school. A mere 10 hours of specific training on this video game included within the initial phase of introductory flight school in the Israeli Air Force was sufficient to significantly improve scores on pilots' check flights in real airplanes.[31] The training "game" was adopted by the United States Air Force for training helicopter pilots.[32]

Although Space Fortress is highly correlated with more sophisticated flight task performance,[31] not all video games are created equal. Performance on select video games correlates with surgical performance, while performance on other video games does not,[33–35] although it is not clear which aspects of the games' design influence their training effectiveness.[36] Similarly, there are examples of high-tech medical virtual reality simulators with excellent face validity that show negative transfer to the targeted skills.[37]

One set of training and assessment tasks that exemplifies good use of task-analytic techniques can be found in the Fundamentals of Laparoscopic Surgery (FLS). Seven tasks were initially examined through validation research to determine which tasks correlated with expertise in laparoscopic surgery. Five tasks whose performance predicted expertise were selected to be part of a task battery used for FLS certification.[38] The tasks selected involved (1) moving rubber rings from one peg to another on a pegboard, (2) cutting a circle in gauze, (3) placing a ligating loop on a foam-rubber appendage, and (4–5) tying 2 specific types of knots in a specified manner. To become FLS certified, all 5 tasks must be performed in a way that meets specific criteria. These 5 tasks have limited resemblance to direct skills used in the operating room. Because the task requirements are not directly related to

surgical procedures or real techniques, FLS prompts complaints from learners (eg, "I never tie knots this way in the OR"); yet, the underlying skills needed to complete the abstract tasks are the same as those needed to be successful as a laparoscopic surgeon. The tasks require the ability to coordinate surgical instrument motion simultaneously using both hands, to judge 3-dimensional depth from a 2-dimensional video display, to demonstrate dexterity with both right and left hand independently, as well as other abilities. These task characteristics are the same ones required for laparoscopic surgery. In fact, performance on the abstract tasks predicts performance in the OR, and receiving training on these tasks improves evaluations of performance in the OR.[39]

Abstract tasks may be more appropriate than complex, realistic tasks in shaping the proper performance strategies in early skill acquisition. Task design should be driven by accurate representations of a task that are at a level of detail and fidelity that is appropriate for the learners and the learning goals.[21,40] However, designing tasks for training can be a challenge. Fully describing the many relevant task-analytic techniques themselves is beyond the scope of this article. What is within scope is to raise awareness of the concept of task structure and of training approaches that can address abstract, broad-based skills without having face validity.

Some simple strategies can facilitate this proposed shift in thinking about simulation. When designing a training task for complex skills such as surgical task simulation, consider using an abstraction of the task. Consider face validity only after other validity and fidelity questions are settled, thus balancing face validity with underlying structure. Using tasks with sound underlying semantic fidelity may require an initial investment in task analysis and task validation. Yet the payoff can be great. Lower-cost alternatives to expensive training modalities can be discovered, leading to more efficient training and greater access to training resources at lower cost.[41]

Conclusion

The HF techniques described above have been broadly applied in aviation and industry for years. However, their adoption within undergraduate and graduate medical education and CME has been sparse. The time-tested techniques of bandwidth feedback, dual-task measures, and task analysis may provide solutions to challenges facing educators, facilitating better assessment and training.

The three example techniques described represent a small sample of the broad range of human factors techniques. HF is an interdisciplinary domain that delves into topics such as safety science, reliability engineering, technology design, and organizational change—all of which are relevant to systems-based practices. Additionally, HF professionals

TABLE 1. Moore's Framework[42] for Assessing Learners and Evaluating Instructional Activities

Outcomes framework	Description	Data sources and methods
Participation LEVEL 1	Number of learners who participate in the educational activity	Attendance records
Satisfaction LEVEL 2	Degree to which expectations of participants were met regarding the setting and delivery of the educational activity	Questionnaires/surveys completed by attendees after an educational activity
Learning: Declarative Knowledge LEVEL 3A	The degree to which participants state what the educational activity intended them to know	Objective: Pre- and posttests of knowledge Subjective: Self-report of knowledge gain
Learning: Procedural Knowledge LEVEL 3B	The degree to which participants state how to do what the educational activity intended them to know how to do	Objective: Pre- and posttests of knowledge Subjective: Self-reported gain in knowledge (eg, reflective journal)
Competence LEVEL 4	The degree to which participants show in an educational setting how to do what the educational activity intended them to be able to do	Objective: Observation in educational setting (eg, online peer assessment and EHR chart stimulated recall) Subjective: Self-report of competence; intention to change
Performance LEVEL 5	The degree to which participants do what the educational activity intended them to be able to do in their practices	Objective: Observed performance in clinical setting; patient charts; administrative databases Subjective: Self-report of performance
Patient health LEVEL 6	The degree to which the health status of patients improves due to changes in the practice behavior of participants	Objective: Health status measures recorded in patient charts or administrative databases Subjective: Patient self-report of health status
Community health LEVEL 7	The degree to which the health status of a community of patients changes due to changes in the practice behavior of participants	Objective: Epidemiological data and reports Subjective: Community self-report

often have expertise in the design, development, and testing of performance measurement tools, tests, surveys, and questionnaires, as well. HF professionals can be valuable partners for collaboration in continuing professional development, and often possess skill sets and analytic tools that compliment those of clinical educators wishing to expand their educational practices.

HF can be applied to assessments of outcomes at multiple levels. For example, in examining Moore's framework for assessing learners and evaluating instructional activities,[42] it becomes evident that HF interventions can be effectively applied to nearly all levels of the framework (TABLE 1). There are HF interventions that can be applied to designing educational content and assessment (Level 2, satisfaction) in a number of ways. For example, as mentioned earlier, the subjective task load index of the National Aeronautics and Space Administration (NASA)[12] combines subjective assessment with a more advanced score-weighting algorithm for

creating a valid and reliable workload assessment. Simple measures of satisfaction could be augmented through a similar weighting technique to produce a more nuanced measure that takes into account the different aspects of satisfaction, and combines them into a reliable, quantitative index. There are psychometric testing techniques for testing declarative knowledge (Level 3), and assessing skill efficiently (Level 4).[43] Detailed motion capture techniques can be used to analyze expert surgical performance (Level 5).[44]

The three techniques offered in this paper focus on bridging the gap between Level 4 (L4) and Level 5 (L5), transitioning from the classroom to the real world. Measuring performance in this transition can be difficult, as it is a continual challenge to create experiences in a controlled environment that will effectively bring about improvement in the clinical setting—transitioning from proctor readiness, where a clinician can perform under supervision, to physician competence, where they can perform independently.

Lessons for Practice

- When designing a training task for complex skills such as surgical task simulation, consider using an abstraction of the task.

- Visual fidelity is only one type of fidelity in simulation. Consider not only physical, but also semantic and phenomenological fidelity, as well.

- Consider face validity only after other validity and fidelity questions are settled. Balance face validity with underlying structure of the task of interest.

- When measurement of performance on a task is difficult because of a "ceiling" effect in measures such as speed and accuracy, using a dual-task paradigm can create a measure that is sensitive to expertise level.

HF provides techniques to address this difficulty. The dual-task paradigm can help differentiate between performance that appears competent in the educational setting but may yet be lacking when attempted in a clinical setting. Bandwidth feedback provides a way to transition between L4 and L5 by creating an environment in training that more closely matches the clinical setting, and supports development of performance strategies that are effective there. Examining face validity and underlying structure can provide insight into the essential elements of clinical work that must be presented in training environment.

Deploying these techniques may require significant study or expertise. Some will require in-depth analysis or investment of resources, such as task analyses related to the structure of tasks. Implementation of these techniques may require additional assistance of a more experienced HF personnel, who can provide valuable collaboration in such endeavors. Although they may require additional effort to employ, the concepts are time tested and proven. They could help to overcome challenges frequently faced in the education of physicians and surgeons.

References

1. International Ergonomics Association Executive Council. IEA defininition of ergonomics. In: Karwowski W, ed. *International Encyclopedia of Ergonomics and Human Factors*. New York, NY: Taylor & Francis; 2000.

2. Gilbreth FB, Carey EG. *Cheaper by the Dozen*. New York, NY: T. Y. Crowell Co; 1948.

3. Apgar V. A proposal for a new method of evaluation of the newborn infant. Curr *Res Anesth Analg.* 1953;32(4):260–267.

4. Roscoe SN. Incremental transfer effectiveness. *Hum Factors.* 1971;13(6):561–567.

5. Skinner BF. An operant analysis of problem solving. *Behav Brain Sci.* 1984;7(4):583–591.

6. Lintern G. Transfer of landing skill after training with supplementary visual cues. *Hum Factors.* 1980;22(1):81–88.

7. Lintern G, Roscoe SN, Sivier JE. Display principles, control dynamics and environmental factors in pilot training and transfer. *Hum Factors.* 1990;32(3):299–317.

8. de Groot S, de Winter JC, López García JM, et al. The effect of concurrent bandwidth feedback on learning the lane-keeping task in a driving simulator. *Hum Factors.* 2011;53(1):50–62.

9. Lintern G. An informational perspective on skill transfer in human-machine systems. *Hum Factors.* 1991;33(3):251–266.

10. Kahneman D, Slovic P, Tversky A. *Judgment Under Uncertainty: Heuristics and Biases.* New York, NY: Cambridge University Press; 1982.

11. Hankins TC, Wilson GF. A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviat Space Environ Med.* 1998;69(4):360–367.

12. Hancock PA, Meshkati N. *Advances in Psychology, Vol. 52: Human Mental Workload.* New York, NY: Elsevier Science; 1988.

13. Wickens CD. Processing resources and attention. In: Damos DL, ed. *Multiple Task Performance.* London, England: Taylor & Francis; 1991:3–34.

14. Gentner DR. The acquisition of typewriting skill. *Acta Psychologica.* 1983;54(1–3):233–248.

15. Shaffer LH. Multiple attention in continuous verbal tasks. In Rabbitt PMA, Dornic S, eds. *Attention and Performance V.* New York, NY: Academic Press; 1975:157–167.

16. Pashler H. Dual-task interference in simple tasks: data and theory. *Psychol Bull.* 1994;116(2):220–244.

17. Grondin S. Timing and time perception: a review of recent behavioral and neuroscience findings and theoretical directions. *Atten Percept Psychophys.* 2010;72(3):561–582.

18. Weinger MB, Reddy SB, Slagle JM. Multiple measures of anesthesia workload during teaching and nonteaching cases. *Anesth Analg.* 2004;98(5):1419–1425.

19. Stefanidis D, Scerbo MW, Montero PN, et al. Simulator training to automaticity leads to improved skill transfer compared with traditional proficiency-based training: a randomized controlled trial. *Ann Surg.* 2012;255(1):30–37.

20. O'Donnell RD, Eggemeier FT. Workload assessment methodology. In: Boff KR, Kaufman L, Thomas JP, eds. *Handbook of Perception and Human Performance, Vol. 2: Cognitive Processes and Performance.* New York, NY: John Wiley & Sons; 1986:42-1–42-49.

21. Dieckmann P, Gaba D, Rall M. Deepening the theoretical foundations of patient simulation as social practice. *Simul Healthc.* 2007;2(3):183–193.

22. Militello LG, Hutton RJB. Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. *Ergonomics.* 1998;41(11):1618–1641.

23. Kirwan B, Ainsworth L, eds. *A Guide to Task Analysis.* London, England: Taylor & Francis; 1992,

24. Wickens CD. Multiple resources and mental workload. *Hum. Factors.* 2008;50(3):449–455.

25. Simon HA. *The Sciences of the Artificial.* Cambridge, MA: MIT Press; 1981.

26. Vincente KJ, Rasmussen J. Ecological interface design: theoretical foundations. *IEEE Trans Syst Man Cybern.* 1992;22(4):589–606.

27. Vicente KJ. Work domain analysis and task analysis: A difference that matters. In: Schraagen JM, Chipman SF, Shalin VL, eds. *Cognitive Task Analysis*. Mahwah, NJ: Lawrence Erlbaum; 2000:101–118.

28. Roscoe SN. Simulator qualification: just as phony as it can be. *Int J Aviat Psychol*. 1991;1(4):335–339.

29. Salkini MW, Doarn CR, Kiehl N, et al. The role of haptic feedback in laparoscopic training using the LapMentor II. *J Endourol*. 2010;24(1):99–102.

30. Mané AM, Donchin E. The Space Fortress game. *Acta Psychologica*. 1989;71:17–22.

31. Gopher D, Weil M, Bareket T. Transfer of skill from a computer game trainer to flight. *Hum Factors*. 1994;36:387–405.

32. Hart SG, Battiste V. Flight test of a video game trainer. In: *Proceedings of the Human Factors Society 26th Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society; 1992:1291–1295.

33. Rosser JC Jr, Lynch PJ, Cuddihy L, et al. The impact of video games on training surgeons in the 21st century. *Arch Surg*. 2007;142(2):181–186; discusssion 186.

34. Bokhari R, Bollman-McGregor J, Kahoi K, et al. Design, development, and validation of a take-home simulator for fundamental laparoscopic skills: using Nintendo Wii for surgical training. *Am Surg*. 2010;76(6):583–586.

35. Kolga Schlickum M, Hedman L, Enochsson L, et al. Transfer of systematic computer game training in surgical novices on performance in virtual reality image guided surgical simulators. *Stud Health Technol Inform*. 2008;132:210–215.

36. Lynch J, Aughwane P, Hammond TM. Video games and surgical ability: a literature review. *J Surg Educ*. 2010;67(3):184–189.

37. Muresan C 3rd, Lee TH, Seagull J, Park AE. et al. Transfer of training in the development of intracorporeal suturing skill in medical student novices: a prospective randomized trial. *Am J Surg*. 2010;200(4):537–541.

38. Peters JH, Fried GM, Swanstrom LL, et al. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery*. 2004;135(1):21–27.

39. Sroka G, Feldman LS, Vassiliou MC, et al. Fundamentals of laparoscopic surgery simulator training to proficiency improves laparoscopic performance in the operating room-a randomized controlled trial. *Am J Surg*. 2010;199(1):115–120.

40. Horton W. Mix media, not metaphors. *Technical Communications*. 1994;41(4):781–783.

41. Seagull FJ, George I, Ghaderi I, et al. Surgical Abdominal Wall (SAW): a novel simulator for training in ventral hernia repair. *Surg Innov*. 2009;16(4):330–336.

42. Moore DE Jr, Green JS, Gallis HA. Achieving desired results and improved outcomes: integrating planning and assessment throughout learning activities. *J Contin Educ Health Prof*. 2009;29(1):1–15.

43. Brannick MT, Prince A, Prince C, Salas E. et al. The measurement of team process. *Hum Factors*. 1995;37(3):641–651.

44. Saleh GM, Gauba V, Sim D, et al. Motion analysis as a tool for the evaluation of oculoplastic surgical skill: evaluation of oculoplastic surgical skill. *Arch Ophthalmol*. 2008;126(2):213–216.