# Alternative Markers of Performance in Simulation: Where We Are and Where We Need To Go

Ann M. Willemsen-Dunlap, PhD, CRNA, Emily S. Binstadt, MD, MPH, Michael C. Nguyen, MD, Nicole C. Elliott, DO, Alan R. Cheney, MD, Ronald H. Stevens, PhD, and Suzanne Dooley-Hash, MD

## ABSTRACT
This article on alternative markers of performance in simulation is the product of a session held during the 2017 *Academic Emergency Medicine* Consensus Conference "Catalyzing System Change Through Health Care Simulation: Systems, Competency, and Outcomes." There is a dearth of research on the use of performance markers other than checklists, holistic ratings, and behaviorally anchored rating scales in the simulation environment. Through literature review, group discussion, and consultation with experts prior to the conference, the working group defined five topics for discussion: 1) establishing a working definition for alternative markers of performance, 2) defining goals for using alternative performance markers, 3) implications for measurement when using alternative markers, identifying practical concerns related to the use of alternative performance markers, and 5) identifying potential for alternative markers of performance to validate simulation scenarios. Five research propositions also emerged and are summarized.

Conventional performance markers include observed behaviors captured by simple checklists and behaviorally anchored rating scales, individual and team self-assessment, data collected automatically by the simulation system, narrative field notes, and comprehensive portfolios of learner performance curated over time. Each of these assessment types has associated performance markers that are well defined; however, they often lack granularity, which limits their ability to offer tangible recommendations for performance.[1] The growth in sensor technology and information processing tools offer the potential for alternative performance markers to address these issues and:

- Provide a detailed scientific description of how people learn (and forget) and how social coordination emerges from the interactions of diverse individuals with and within a complex changing environment.[2]
- May provide new insights about ways in which cognition supports decision making among clinicians with all levels of experience.

## CONSENSUS AREAS DISCUSSED

The breakout group discussed five areas concerning alternative markers of performance. They are summarized below.

### Working Definition and Examples of Alternative Markers of Performance

Conventional performance markers, including expert observation, typically generate high-level data that views an individual or a team as a system interacting with the environment. Such markers contribute to the understanding of large-scale (i.e., longer-term) patterns

From Interprofessional Education, Jump Simulation Center (AMWD), Peoria, IL; the Department of Emergency Medicine, Regions Hospital (ESB), St. Paul, MN; the Department of Emergency Medicine, Lehigh Valley Healthcare System (MCN, NCE), Allentown, PA; the Department of Emergency Medicine, Thomas Jefferson University (ARC), Philadelphia, PA; UCLA (emeritus) and The Learning Chameleon (RHS), Culver City, CA; and the Department of Emergency Medicine, University of Michigan (SDH), Ann Arbor, MI.
Received June 17, 2017; revision received September 19, 2017; accepted September 20, 2017.
Supervising Editor: Rosemarie Fernandez, MD.
Address for correspondence and reprints: Ann M. Willemsen-Dunlap, PhD, CRNA; e-mail: ann.m.willemsen-dunlap@osfhealthcare.org
ACADEMIC EMERGENCY MEDICINE 2018;25:250–254.


ISSN 1553-2712

doi: 10.1111/acem.13321

and trends.[3,4] Intermediate markers such as communication analysis generate data that may bridge and validate both micro (milliseconds to seconds) and macro (tens of minutes to days) level performance data.[5] Markers that generate micro-level data contribute to the understanding of subsystems (such as those in the brain that underpin performance) tend to be highly granular. An example of such data is modeled electroencephalography (EEG) data. Sampled at millisecond intervals, EEG-generated data provide a window into the microevents, e.g., neuronal firing, in the brain that underpin learners' responses and understanding or lack thereof.[6] Such data may provide a more targeted approach to training each level of performance and offer the potential to objectively quantify parameters of performance among individuals and teams.[7]

Conventional performance markers include observed behaviors captured by simple checklists and behaviorally anchored rating scales, individual and team self-assessment, data collected automatically by the simulation system, narrative field notes, and comprehensive portfolios of learner performance curated over time.[1,2] Each of these assessment types has associated performance markers that are well defined; however, they often lack granularity, which limits their ability to offer tangible recommendations for performance.[4] The growth in sensor technology and information processing tools may offer the potential for alternative performance markers to address these issues and

- Provide a detailed scientific description of how people learn (and forget) and how social coordination emerges from the interactions of diverse individuals with and within a complex changing environment.
- May provide new insights about ways in which cognition supports decision making among clinicians with all levels of experience.

## Working Definition of Alternative Performance Markers

A broad working definition of alternative markers proposed at the consensus conference was, "a performance marker that can potentially or is likely to contribute benefit, but whose infrastructure, either in material or personnel, is not yet present to make it practical." Working group and breakout session members refined this definition with the following characteristics. It is important to note that they will not be common to all alternative markers (Table 1).

Alternative performance markers are generated from various types of data. Because data sources that generate alternative performance markers are either in immediate contact with the body (on-body)[6,8–12] or are not in immediate contact with the body (off-body),[5,6,11–13] the sources are presented for clarity in Table 2 as on-body or off-body. A description of this data is also included in Table 2.

## Identification of Goals for Using Alternative Performance Markers

Breakout session members identified as important the goals of using alternative markers of performance to develop research-based, quantitative answers to

- Elucidate learning processes and the development of long- and short-term memory during clinical tasks;
- Understand the cognitive processes that support team cohesion and coordination;

Table 1
Common Alternative Marker Attributes

| | |
|---|---|
| Generates granular data | Granular data are broken down into the smallest possible pieces to generate detail. Granular data can be modeled in any way the scientist requires. It is possible to aggregate and disaggregate such data to meet needs of different situations. |
| Continuous nature of data | Data are captured in uninterrupted fashion during an assessment session. |
| Automated data collection | Preestablished protocols drive computerized data collection from on- and off-body sensors. |
| Generates large quantities of data | Ever-growing array of sensors with high sampling rates will generate multiple measurements from each sample from a data source. |
| Raw signals requiring processing and modeling | EEG, fNIRS, examples of raw signals that must be processed into data and then mathematically modeled. |
| Available as individual and/or team data | Some alternative markers hold potential to untangle individual's contribution to team performance. |
| Near real time | Will likely approach the ability to process signals and model alternative marker data in near real time. |

EEG = electroencephalography; fNIRS = functional near infrared spectroscopy.

Table 2
On-body and Off-body Data Sources for Alternative Performance Markers

| Data Source | Description |
|---|---|
| Off-body | |
| Computerized communication analysis[5] | Communication characteristics linked to specific processes and team performance. |
| Galvanic skin response and vocal stress cues[17] | Synchronized autonomic arousal as measured by changes in skin conductance and elements of speech including pitch, rate, and loudness. |
| Oculometrics[16,17] | Evaluates pupil size to measure autonomic arousal. |
| Eye tracking[17] | Measures either the point of gaze or the motion of an eye relative to the head. |
| Audiovisual data analysis driven by machine learning[18] | Example applications include large-scale analysis of discourse, actions, gestures, tone of voice, and other body language captured via AV recording; driven by machine learning. |
| On-body | |
| EEG[4] | Measures the electrophysiology of action potentials within the brain; does so across multiple frequencies. |
| fMRI[9] | Measures activity in different parts of the brain by evaluating oxygen levels in the blood circulating there. |
| fNIRS[10] | Use of NIRS to measure hemodynamic changes in the brain that are associated with neuronal behavior. |
| Electrocardiogram for HRV[17] | HRV refers to normal variation in time between heartbeats; used as a marker of autonomic arousal. |
| Cortisol, interleukin, neuropeptide Y, interferon-$\gamma$, tumor necrosis factor[17] | Biochemical markers of autonomic arousal and stress. |

EEG = electroencephalogram; fMRI = functional MRI; fNIRS = functional near infrared spectroscopy; HRV = heart rate variability.

- Provide objective metrics to evaluate the efficacy of simulation-based curricula;
- Support real-time training adjustments and feedback to maximize learning;
- Further describe the cognitive processes supporting decision-making and provide insight into these processes for the learner.

## Implications for Measurement

The introduction of alternative performance markers raised several questions around measurement. The first was whether or not traditional theories of validity such as those introduced by Messick[14] and Kane[15] would remain relevant when analyzing data from alternative performance markers. There was broad consensus that these constructs would remain central to measurement. Participants also agreed that multimodal approaches to validation of alternative markers would be important and that such studies should include intermediate markers of performance such as speech analysis that can bridge micro-events such as neuronal firing and macro-level behavioral observations done by trained and calibrated expert raters. Preliminary results suggest that this multimodal approach may have utility in situations as diverse as submarine navigation tasks by bridge crews and teamwork in health care.[5,6] Multimodal approaches may also make it possible to more routinely provide the simulation and education communities with quantitative descriptions of the relationship between team members with each other, with complex changing environments and across time and task sets.[16]

## Practical Concerns Related to the Use of Alternative Performance Markers

The conference attendees discussed several practical concerns related to alternative markers including cost, infrastructure, data handling, and end-user acceptance. Regarding cost and infrastructure, many alternative markers will require an investment in new sensors as well as computing and other processing equipment to collect and prepare data, then analyze and integrate the results into meaningful output. One could imagine a fully equipped simulation-based performance laboratory to gather and analyze off-body and on-body performance markers such those listed in Table 2. The price tag on such a facility would be substantial and likely out of reach for many simulation programs in the beginning. It was recognized that making rational decisions about which technologies to invest in would require deliberate and far-ranging conversations among multiple stakeholders, including department administrators, educational leaders and researchers, and others.

**Sensors, Processing, Integration, and Use of Data.** Alternative markers are expected to generate large quantities of data, especially as improvements are made in sensor technology and computer algorithms. The large quantities of data generated by alternative markers creates the need to be able to record, process, integrate, and visualize data in meaningful ways. Researchers need to develop methods and analytic approaches to this "big data" problem keeping in mind critical issues related to level(s) of analysis.

**Acceptance of Alternative Markers of Performance.** Research and education-focused conference attendees noted that acceptance of alternative performance markers by the emergency medicine simulation community could represent a significant barrier. Training programs have traditionally tried to move learners along a predetermined path toward competency. However, with alternative marker data, the potential for real-time assessment and feedback offers the opportunity for rapid adjustments in training design and implementation. Such an approach would require a paradigm shift in clinical education. Educators would need to master the use of alternative marker data to guide rapid adaption of learning goals, objectives, and delivery of the simulation to learners. Likewise, learners would need to be prepared for a more dynamic, individualized curricula.

### Potential for Alternative Markers of Performance to Validate Simulation Scenarios

Alternative marker data can help educators and learners alike focus on scenario elements that are most important for reaching training objectives. For example, educators may wish to design a scenario that requires specific cognitive functions. Alternative markers can provide data corroborating the activation of cognitive processes when expected in the scenario. Research will be needed to evaluate the benefit of using alternative performance markers to understand more deeply the efficacy of various simulation modalities for different learning needs.

## AREAS FOR FUTURE RESEARCH

The following research propositions emerged from the consensus conference breakout session:

1. should focus on providing validity evidence to support the use of alternative markers in both individual and team-based performance assessments.
2. should consider collecting alternative marker data in actual clinical environments to facilitate the evaluation of system and process changes on performance.
3. is needed to determine appropriate methodologic and statistical approaches to alternative marker data aggregation and presentation.
4. Educators need further instruction to support effective incorporation of alternative marker data into simulation-based training design and implementation.
5. Research evaluating the effectiveness of simulation-based training should incorporate alternative marker data when appropriate.

## SUMMARY

Alternative performance markers hold significant promise for quantitating performance at a level of biobehavioral detail never before realized. As these markers move from leading-edge research to common use, it is incumbent on the simulation and assessment communities to actively participate in discussions and research necessary to establish best practices for collection, analysis, and use of data from alternative markers. These best practices must rest on a firm foundation of science drawn from biologic, computational, computer, measurement, and behavioral realms. With such a foundation to support their development, deployment, and use, today's alternative performance markers may become tomorrow's conventional measures.

The authors acknowledge Donald Halpin, Jump Simulation; Hannah Bond, Tulane University

## References

1. Marriage B, Kinnear J. Assessing team performance – markers and methods. Trends Anaesth Crit Care 2016;7–8:11–6.
2. Tognoli E, Kelso JA. The coordination dynamics of social neuromarkers. Front Hum Neurosci 2015;9:563.
3. Jackson A, Marks LB, Bentzen SM, et al. The lessons of QUANTEC: recommendations for reporting and gathering data on dose–volume dependencies of treatment outcome. Int J Radiat Oncol Biol Phys 2010;76:S155–60.
4. Stevens R, Galloway T, Halpin D, Willemsen-Dunlap A. Healthcare teams neurodynamically reorganize when resolving uncertainty. Entropy 2016;18:427.
5. Gorman JC, Martin MJ, Dunbar TA, et al. Cross-level effects between neurophysiology and communication during team training. Hum Factors 2015;58:181–99.

6. Stevens RH, Galloway T. Are neurodynamic organizations a fundamental property of teamwork? Front Psychol 2017;8:644.
7. Stevens R, Galloway T, Lamb J, Steed R, Lamb C. Linking team neurodynamic organizations with observational ratings of team performance. In: Innovative Assessment of Collaboration. Cham: Springer International Publishing; 2017. p. 315–30.
8. Erk S, Kleczar A, Walter H. Valence-specific regulation effects in a working memory task with emotional context. Neuroimage 2007;37:623–32.
9. Howard SJ, Burianová H, Ehrich J, et al. Behavioral and fMRI evidence of the differing cognitive load of domain-specific assessments. Neuroscience 2015;297:38–46.
10. Dan M, Saha A, Konar A, Ralescu AL, Nagar AK. A type-2 fuzzy approach towards cognitive load detection using fNIRS signals. 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). Piscataway (NJ): IEEE; 2016. p. 2508–15.
11. Fishburn FA, Norr ME, Medvedev AV, Vaidya CJ. Sensitivity of fNIRS to cognitive state and load. Front Hum Neurosci 2014;8:76.
12. Guastello SJ, Reiter K, Malon M, et al. Cognitive workload and fatigue in a vigilance dual task: miss errors, false alarms, and the impact of wearing biometric sensors while working. Nonlinear Dynamics Psychol Life Sci 2016;20:509–35.
13. Cook DA. Technology enhanced simulation to assess health professionals. Acad Med 2013;88:872–83.
14. Validity Messick S. In: Linn RL, editor. The American Council on Education/Macmillan Series on Higher Education Educational Measurement. New York: Macmillan, 1989:13–103.
15. Kane M. An argument based approach to validity. Psychol Bull 1992;112:527–35.
16. Salas E, Stevens R, Gorman J, Cooke NJ, Guastello SJ, von Davier A. What will quantitative measures of teamwork look like in 10 years? Proc Hum Factors Ergon Soc Annu Meet 2015;59:235–9.
17. Galster SM, Johnson EM. Sense-assess-augment: a taxonomy for human effectiveness. (Report AFRL-RH-WP-TM-2013-0002). Wright-Patterson Air Force Base (OH): US Air Force Research Laboratory; 2013.
18. Microsoft Solution Providers. Microsoft Data Mining. Amsterdam: Elsevier, 2001:255–87.