Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting - 2015

WHAT WILL QUANTITATIVE MEASURES OF TEAMWORK LOOK LIKE IN 10 YEARS?

Session Chair: Eduardo Salas, Ph.D.
Rice University, Houston, TX

Panelists:

Ron Stevens, Ph.D., UCLA School of Medicine, The Learning Chameleon, Inc., Los Angeles, CA
Jamie Gorman, Ph.D., Georgia Institute of Technology, Atlanta, GA
Nancy J. Cooke, Ph.D., Arizona State University, Mesa, AZ
Stephen Guastello, Ph.D., Marquette University, Milwaukee, WI
Alina A. von Davier, Ph.D., Educational Testing Service, Inc., Princeton, NJ

INTRODUCTION

Nationwide there is a need for cost-effective training solutions that are highly automated, adaptable, and capable of producing quantifiable behavioral changes in teams that are indicative of deep learning. In contrast to what is known about individual skill acquisition and persistence, relatively little is known about how team process skills develop; how well these skills once learned in one context transfer to another context; how long the skills persist when unused; and, what interventions or training will most rapidly restore them?

Answering these questions is challenging due to the limited number of quantitative teamwork measures that track team performance, cohesion and flexibility across teams, time, environments and training protocols. Adopting a scientific approach for studying the effects of training interventions is problematic without theory and methods that are aligned and capable of representing and capturing the dynamics of team performance.

A confluence of new technologies will soon generate enormous amounts of new data at an unprecedented level of detail about teams. But these data will also raise questions of their own; principally how researchers will make sense of the expected onslaught of data and derive general organizing principles that guide the co-evolution of the complex team and task interactions. This suggests the need for novel methods and ways of thinking about team dynamics and measurement.

Our goals are to speculate, given where we are, where the measurement and assessment of learning and performance of teams and of individuals in teams might go in the next decade and how we might get there. As such, some sample 'Big' Questions' that the panelists were asked to consider in their presentations include:

- What types of high and low level data abstractions might provide the most useful quantitative information about teamwork?
- Across which biologic and interpersonal scales of teamwork will the strongest information flows be found?
- Can dynamical clues tell us how well a team is performing / will perform?
- In addition to performance assessment, what can we learn from dynamics about team flexibility, cohesion, leadership, and resilience?
- How can we disentangle individual contributions from the team contribution and accurately measure them?

PANELIST ABSTRACTS

Toward Quantitative Descriptions of the Neurodynamic Organizations of Teams

Ron Stevens
UCLA School of Medicine,
The Learning Chameleon, Inc.

Advances in our understanding of the learning process used by teams while they develop and refine their team skills have been slowed by a lack of easily understandable quantitative approaches that can objectively and automatically assess collective learning processes and outcomes over time in training situations.

We have developed symbolic models of teamwork that capture the brainwave levels of each person of the team, and situates them in the context of the levels of other team members as well as the immediate context of the task. Quantitative estimates of the symbol variations in the data stream are then made using a moving window of entropy approach. Periods of decreased entropy represent times of increased team neurodynamic organization when there were prolonged and restricted neurodynamic relationships across members of the teams.

Using this approach we have shown that (a) the neurodynamic rhythms of six-person US Navy submarine navigation teams are measurable and are entrained by the task (Stevens, Gorman, Amazeen, Likens & Galloway, 2013); (b) the structure of these rhythms is multifractal, resulting from the meso and micro responses of teams to changes in the task and the sharing of information across the crew (Likens, Amazeen, Stevens, Galloway & Gorman, 2014); and (c) quantitative differences in team's neurodynamic rhythms are linked with team expertise, resilience, and communication

Copyright 2015 Human Factors and Ergonomics Society. DOI 10.1177/1541931215591048

(Stevens & Galloway, 2014). Consistent with the nonlinear dynamical systems concepts of elasticity and rigidity in complex adaptive systems, the expert navigation teams were positioned at an organizational neurodynamic point midway between rigid and elastic. These findings suggest the existence of team-related neurodynamical processes that quantitatively track across the novice-expert continuum.

Team Performance at the Level of Emergent, Dynamic Coordinative Relations

Jamie C. Gorman
Georgia Institute of Technology

In a recent series of experiments on dyadic, interpersonal coordination, we found that the tendency to spontaneously synchronize one's movements with those of a teammate can interfere with human performance in team tasks like assisted suturing and knot tying (Gorman & Crites, 2014). Namely, we found that the ability to decouple the hands, such that they move independently (i.e., to *not* synchronize), is fundamental to tying skill and is presumably acquired from an early age. In the context of unfamiliar, dyadic tying, however, participants were unable to fight the spontaneous tendency to synchronize their hand movements, which hurt team performance. How do we account for phenomena such as this in the science of team learning? This example is meant to illustrate the need for team skill acquisition and assessment to account for high-level, emergent coordinative relations, such as "sync" (Strogatz, 2003) that structure team performance beyond the control of the individual.

Interpersonal tying provides a relatively simple example of how emergent, dynamic coordinative relations structure team performance, but we see the same phenomena at work in more complex, cognitive settings, such as intelligence analysis and planning and team command-and-control. In those situations, team members self-organize "cognitive-behavioral" (e.g., communication) patterns without being consciously aware of it (Dunbar & Gorman, 2014). We think the key to understanding how emergent coordinative relations shape team performance is in identifying the mathematical and statistical dynamics (e.g., sync; self-organization) that occur as team members interact. In the next decade, this approach will not produce new models of shared cognition but will be characterized by mathematical models that go beyond the individual mental state or top-down knowledge with the goal of understanding how general coordinative mechanisms structure team performance. Such models will also have implications for understanding how individual psychological processes are structured by emergent, coordinative relations (Gorman, 2014).

This dynamic approach to understanding team performance has already provided novel predictions and ways of thinking about enhancing team performance, including enhancing team flexibility and resilience. I will briefly describe research conducted using the dynamic perspective that has already provided new insights into how teams develop, what develops, and how to enhance transfer to novel contexts during skill acquisition in both motor-perceptual and cognitive-behavioral tasks. I will link these results together by briefly describing the common theoretical core that underlies them, and I will briefly describe the types of models and methodologies that are needed to understand team performance at the level of emergent, dynamic coordinative relations.

Communication Dynamics for Team Assessment

Nancy J. Cooke
Arizona State University

Teams can be viewed as complex dynamic systems made up of interacting components that are systems themselves. As effective teams learn, not only does their performance improve, but their team process behaviors evolve to become more flexible, adaptive, and resilient. These process dynamics have been linked to team effectiveness (Cooke, Gorman, Myers, & Duran, 2013).Therefore assessment of team learning can benefit from an understanding of the dynamics of team interactions. Team interactions often take place through communication, though there are other forms of interaction that are nonverbal including gestures, facial expressions, and implicit interaction. Of these data sources, communication is the most straightforward to collect and therefore, most commonly collected. Communication data can be collected unobtrusively. There are also opportunities to analyze communication data in near-real time for continuous monitoring of team learning and rapid intervention. For instance, metrics based on communication flow from person to person or amount of communication, are amenable to this real-time processing (Cooke & Gorman, 2009). Research is needed on the association between communication dynamics and effective team process and performance in various contexts. Ultimately the discovery of communication dynamics signatures linked to specific process and performance across contexts would allow for impactful and timely interventions.

For example, recent studies in my lab of human-synthetic teammate teams in the unmanned aerial system context suggest that coordination can be impacted by a single team member who knows the ideal communication push and pull. Future team training may benefit from synthetic teammates that are able to serve the role of coordination coach.

Synchronization of Autonomic Arousal in Dyads and Teams

Stephen J. Guastello
Marquette University

Physiological synchronization of autonomic arousal between people is thought to be an important component of work team coordination and other interpersonal dynamics. The group dynamics, in turn, contribute to workload and fatigue effects at the group level in addition to the individually-defined work assignments.

The minimum requirements for two living or non-living entities to synchronize are two oscillators, a feedback loop between them, and a control parameter that speeds up the process (Strogatz, 2003). When speed reaches a critical level, a phase shift occurs such that the system goes into phase-lock. The first challenges for operationalizing these principles to human systems include identifying the nature of the oscillators, the feedback loop, and the control parameter. In the prototype, the feedback is reciprocal between the two oscillating entities. In human systems such as interactions in conversations, both one-way and two-way influences are possible. The coupling is generally loose and moderated by the empathy levels of the two parties (Guastello, Pincus & Gunderson, 2006; Marci, Ham, Moran, & Orr, 2007).

Analytic challenges include: (a) determining the statistical time series models that capture the dynamics of the teams' interaction patterns, (b) finding the optimal lag length between observations for structuring the models, and (c) connecting synchronization parameters to performance variables, subjective ratings of workload and other group dynamics, and individual differences of the contributing group members. Equation 1, for instance, is capable of registering self-organizing and chaotic processes, and it identified the synchronization links in the sample of dyads more often and with greater accuracy compared to a linear alternative. In Eq. 1

$$z_2 = A\,\exp(Bz_1) + \exp(CP_1) \tag{1}$$

z is the normalized behavior (autonomic arousal) of the target person at two successive points in time, P is the normalized behavior of the partner at time 1, and A, B, and C are nonlinear regression weights (Guastello et al., 2006).

Lag length denotes how much real time is required to elapse between the two measurements in order to observe the coupling effect. A measurement at time 2 is a function of itself at time 1 and a coupling effect from another source also at time 1. In a vigilance dual task experiment, 73 undergraduates worked in pairs for 90 min while galvanic skin responses (GSR) were recorded. Event rates on the vigilance task either increased or decreased without warning during the work period. Results based on two criteria supported a lag value of 20 sec (Guastello, Reiter & Malon, in press).

The properties of the linear and nonlinear (Eq.1) autoregressive models, with and without a synchronization component were examined. All models were more accurate at a lag of 20 sec compared to customized lag lengths. Although the linear models were more accurate overall by a margin of 4-13% of variance accounted for, the nonlinear synchronization parameters were more often related to psychological variables and performance. (Guastello, in press). Importantly, greater synchronization was observed with the nonlinear model when the target event rate increased, compared to when it decreased, which was expected from the general theory of synchronization. Equation 1 was also more effective for uncovering inhibitory or dampening relationships between the co-workers as well as mutually excitatory relationships.

The latest study on this theme involves teams of four people who play an emergency response (ER) game against a single opponent, all with GSR recordings. An example data stream appears in Figure 1. The participants appear to be in phase lock. The generalizability of this result remains to be determined. The adaptive value of high levels of synchrony has also been questioned (Stevens, Galloway, & Lamb, 2014).

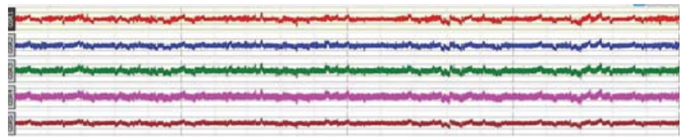

Figure 1. GSR readings for five people in an ER game.

As with many games, the ER team and opponent take turns. It now appears that the optimal lag length is much shorter for this task. The ER team members are synchronized with each other and also with the opponent thus producing the cluster of linkages shown in Figure 2.

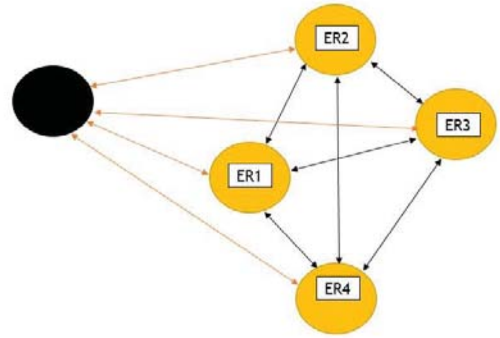

Figure 2. Linkages between ER team members and opponent.

Future research should dissect the experimental tasks to identify the primary oscillators, feedback loops, possible control parameters, and conditions that induce higher levels of synchronization and involve different types of internal group coordination. Analyses of biometrics need to go beyond discrete event-related potentials, which are commonly used, to focus instead on continuous streams of data and the analysis of dynamics therein. The dynamics can then be related to qualitative variables of interest, such as coordination behavior, communication events, workload manipulations, and ratings of group processes, thereby building a comprehensive bio-psychosocial model and generalizable theory.

Collaborative Assessments and Data Analysis

Alina A. von Davier
Educational Testing Service

Educational measurement is undergoing dramatic change at all levels, with new directions in assessment of individuals and groups. Some of the most innovative and exciting features involve conversation-based, technology-enhanced learning and assessment tools in the areas of collaborative assessment, game-based learning, and simulation-based training. Among the advantages of these approaches is that they support learning of cognitive, social and affective skills within a common framework and allow for a detailed collection of the process data in addition to the usual outcome data in structured log files. Collaborative assessments in game-like environments integrate many of these different approaches and tools (Liu, von Davier, Hao, Kyllonen, Zapata-Rivera, 2014).

Collaboration is one of the skills identified as the "21st-century skills" and it receives attention among stakeholders in both higher education and the workplace. The OECD (OECD, 2013) included a test of collaborative problem solving skills in its PISA 2015 survey of critical skills. Collaborative assessment is also being promoted by a global initiative called Assessment and Teaching of 21st Century Skills, a partnership among Cisco, Intel, Microsoft and the University of Melbourne to prepare students to live and work in information-age societies.

The questions for the educational specialists revolve around the measurement issues: how can we measure accurately individual contributions to team success? Shall collaborative tests be domain specific or are the collaborative problem solving skills transferable from one domain to another? How can we integrate the data from the dynamic process of collaboration and the outcome test data and build valid measurement models?

Collaborative interactions in computerized educational environments produce data of extraordinarily high dimensionality (often containing more variables than people for whom those variables are measured). Extracting key features from the noise in such data is crucial not only to make analysis computationally tractable (Masip, Minguillon, & Mor, 2011), but also to extract relevant features of student performance from the noise surrounding them (Kim et al., 2008). Nowadays, with the technological advantages of systems for recording, capturing, and recognition (e.g., Kinect for Windows) of multimodal data, the data from collaborative interactions contain discourse, actions, gestures, tone, body language that result in a deluge of data (See Figure 3). To these types of data we can further add the neurodata collected with (portable) EEG headsets. One way to attempt to find patterns among these different types of data is to make use of data mining techniques.

Data mining does not have a long history in education or psychology because, until recently, educational and psychological data were not often of high enough dimensionality to require such techniques. However, these techniques have been used for decades in fields where data with high dimensionality has long been the norm, such as finance, marketing, medicine, astronomy, physics, chemistry, and computer science (Frawley, Piateski-Shapiro, & Matheus, 1992). The purpose of data mining techniques is to reduce the dimensionality of the dataset something more manageable (Hand, Mannila, & Smyth, 2001) by extracting implicit, interesting, and interpretable patterns (Frawley et al, 1992) in order to allow research questions to be addressed that would not otherwise be feasible (Romero et al., 2011). Data mining methods as those developed by Kerr (in press) and Kerr and Chung (2012) can be applied to identify patterns of interactions and strategies of success in collaborative problem solving tasks.

In one of the recent pilot applications at Educational Testing Service (ETS), we analyzed the behavioral convergence of test takers in dyads in a science collaborative problem solving assessment task, the Tetralogue (Luna-Bazldua, Khan, von Davier, Hao, Liu, Wang, 2015) – see Figure 3. In order to study evidence of behavioral convergence, features from log files and video data of 24 study participants were represented as a multi-dimensional behavioral feature vector composed of cognitive behaviors (such as the *number of messages among the dyad members, the number of requests for help from the system)* and non-cognitive behaviors (such as *engagement, hand-on-face, anxiety, curiosity, anger, joy, contempt and surprise).*

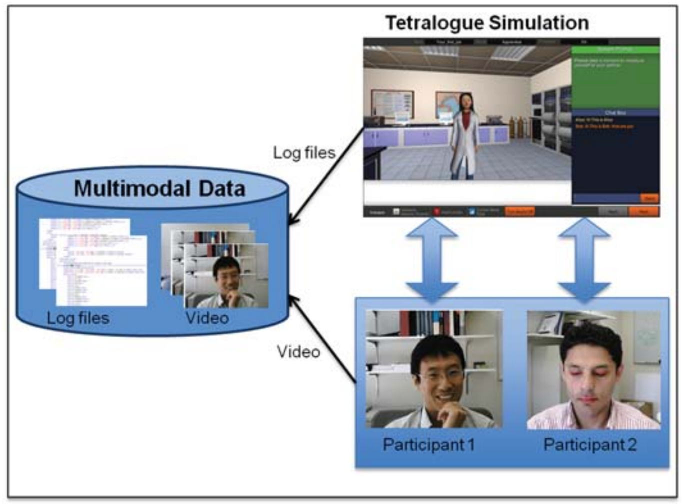

Figure 3. Multimodal data capture including video and action log files while participants solve a problem collaboratively (using the ETS' Tetralogue platform).

A hierarchical cluster analysis was performed on an Euclidean distance matrix (i.e., a similarity matrix) computed from the multidimensional behavioral feature data of the study participants. The cluster analysis revealed that members in the same dyad tended to group together from the beginning of the clustering process (i.e., they will be closer to each other in the feature space than to others). We believe this observed pattern of agglomeration of the dyad partners could be interpreted as evidence of convergence of cognitive and non-cognitive states when people interact in a collaborative task.

Other methods can be considered on these rich data that exhibit time dependence at the individual level are multivariate stochastic processes, time processes, and dynamic models. Some of these models can accommodate analyzing the interactions among multiple team members (von Davier & Halpin, 2013; Halpin & von Davier, 2013); other models are more appropriate for analyzing the states of collaborative events (Soller & Stevens, 2007).

REFERENCES

Cooke, N. J. & Gorman, J. C. (2009). Interaction-Based Measures of Cognitive Systems. *Journal of Cognitive Engineering and Decision Making*: Special Section on: Integrating Cognitive Engineering in the Systems Engineering Process: Opportunities, Challenges and Emerging Approaches *3*, 27-46.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J.L. (2013). Interactive Team Cognition, *Cognitive Science*, 37, 255-285, DOI: 10.1111/cogs.12009.

Dunbar, T. A. & Gorman, J. C. (2014). Fractal effects of task constraints in the self-organization of team communication. Talk presented at *the Human Factors and Ergonomics Society 58th Annual Meeting.*

Frawley, W. J., Piateski-Shapiro, G., & Matheus, C. J. (1992). Knowledge discovery in databases: An overview. *AI Magazine*, *13*(3), 57–70.

Gorman, J. C. (2014). Team coordination and dynamics: Two central issues. *Current Directions in Psychological Science, 23,* 355-360.

Gorman, J. C. & Crites, M. J. (2014). Learning to tie well with others: Bimanual versus intermanual performance of a highly practised skill. *Ergonomics.* Online version published ahead of print.

Guastello, S. J. (in press). Physiological synchronization in a vigilance dual task. *Nonlinear Dynamics, Psychology, and Life Sciences.*

Guastello, S. J., Pincus, D., & Gunderson, P. R. (2006). Electrodermal arousal between participants in a conversation: Nonlinear dynamics for linkage effects. *Nonlinear Dynamics, Psychology, and Life Sciences, 10,* 365-399.

Guastello, S. J., Reiter, K., & Malon, M. (in press). Estimating appropriate lag length for synchronized physiological time series: The electrodermal response. *Nonlinear Dynamics, Psychology, and Life Sciences.*

Halpin, P. F., & von Davier, A. A. (2013). *Evaluating the roles of individual members in team interactions*. Presented at the National Council on Measurement in Education, San Francisco, CA.

Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. Cambridge, Massachusetts: The MIT Press.

Kerr, D. (in press). Using data mining results to improve educational video game design. *Journal of Educational Data Mining*.

Kerr, D., & Chung, G. K. W. K. (2012). Identifying key features of student performance in educational video games and simulations through cluster analysis. *Journal of Educational Data Mining*, *4*, 144–182.

Likens, A., Amazeen, P., Stevens, R., Galloway, T., Gorman, J. (2014). Multifractal scaling reveals the nested nature of team performance. Social Neuroscience. Vol. 9(3), pp. 219-234.

Liu, L., Von Davier, A.A., Hao, J., Kyllonen, P. & Rivera-Zapata, D. (2014). A tough nut to crack: Measuring collaborative problem solving. Paper presented at the *National Council in Measurement in Education Annual Meeting*, Philadelphia.

Luna-Bazldua, D.A., Khan, S., von Davier, A.A., Hao, J., Liu, L., & Wang, Z. (2015). On Convergence of Cognitive and Noncognitive Behavior in Collaborative Activity. Accepted for *Proceedings of the 8th International Conference on Educational Data Mining,* Madrid, Spain

OECD (2013). *PISA 2015 Collaborative Problem Solving Framework*. OECD Publishing. http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf. *Retrieve Dec, 18*, 2014.

Marci, C. D., Ham, J., Moran, E., & Orr, S. P. (2007). Physiologic correlates of perceived therapist empathy and social-emotional process during psychotherapy. *Journal of Nervous and Mental Disease, 195*, 103-111.

Masip, D., Minguillon, J., & Mor, E. (2011). Capturing and analyzing student behavior in a virtual learning environment. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S. J. d. Baker (Eds.), *Handbook of educational data mining* (pp. 339–351). Boca Raton, FL: CRC Press.

Romero, C., Gonzalez, P., Ventura, S., del Jesus, M. J., & Herrera, F. (2009). Evolutionary algorithms for subgroup discovery in e-learning: A practical application using Moodle data. *Expert Systems with Applications*, *39*, 1632–1644.

Soller, A. & Stevens, R. (2007). *Applications of stochastic analyses for collaborative learning and cognitive assessment.* Alexandria: Report for the Institute for Defense Analysis.

Stevens, R. H. & Galloway, T., (2014). Toward a quantitative description of the neurodynamics organizations of teams. *Social Neuroscience* 9:2, 160-173.

Stevens, R.H., Gorman, J.C., Amazeen, P., Likens, A., & Galloway, T. (2013). The organizational neurodynamics of teams. *Nonlinear Dynamics, Psychology and Life Sciences, 17*, No. 1, pp. 67-86.

Stevens, R. H., Galloway, T., & Lamb, C. (2014). Submarine navigation team resilience: Linking EEG and behavioral models. *Proceedings of the Human Factors and Ergonomics Society, 58*, 245-249.

Strogatz, S. (2003). *Sync: The emerging science of spontaneous order*. New York: Hyperion.

Von Davier, A. A., & Halpin, P. F. (2013). *Collaborative problem solving and the assessment of cognitive skills: Psychometric considerations* (No. ETS RR-13-41). Princeton, NJ: Educational Testing Service.

Team Neurodynamics

by The

Learning Chameleon, Inc.