AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Applied Psychology

ISSN: 0021-9010

2022, Vol. 107, No. 4, 533–559
https://doi.org/10.1037/apl0000856

Indexing Dynamic Collective Constructs Using Computer-Aided Text Analysis: Construct Validity Evidence and Illustrations Featuring Team Processes

John E. Mathieu[1], Mikhail A. Wolfson[2], Semin Park[3], Margaret M. Luciano[4], Wendy L. Bedwell-Torres[5], P. Scott Ramsay[6], Elizabeth A. Klock[1], and Scott I. Tannenbaum[7]

[1] Department of Management, University of Connecticut
[2] Department of Management, University of Kentucky
[3] Department of Management and Entrepreneurship, University of Iowa
[4] Department of Management and Organization, Pennsylvania State University
[5] Department of Management, University of Memphis
[6] Department of Psychology, University of South Florida
[7] Group for Organizational Effectiveness

Organizational processes have been widely recognized as both multilevel and dynamic, yet traditional methods of measurements limit our ability to model and understand such phenomena. Featuring a popular model of team processes advanced by Marks et al. (2001), we illustrate a method to use individuals' communications as construct valid unobtrusive measures of collective constructs occurring over time. Thus, the purpose of this investigation is to develop computer-aided text analysis (CATA) techniques that can score members' communications into valid team process measures. We apply a deductive content validity-based method to construct CATA dictionaries for Marks et al.'s dimensions. We then demonstrate their convergent validity with subject matter experts' (SMEs) hand-coded team communications and different SMEs' behaviorally anchored rating scales based on video recordings of team interactions, using multitrait–multimethod analyses in two samples. Using a third sample of paramedics performing a high-fidelity mass casualty incident exercise, we further demonstrate the convergent validity of the CATA and SME scorings of communications. We then model the relationships among processes across episodes using all three samples. Next, we test criterion-related validity using a longitudinal dual-discontinuous change growth modeling design featuring the paramedic CATA-scored team processes as related to a dynamic performance criterion. Finally, we integrate behavioral data from wearable sensor badges to illustrate how CATA can be scored at the individual level and then leveraged to model dynamic networks of team interactions. Implications, limitations, directions for the future research, and guidelines for the application of these techniques to other collective constructs are discussed.

Keywords: team process, CATA, construct validity, multitrait–multimethod

Two important developments have occurred in the field of applied psychology over the past few decades that have increased the complexity of capturing and modeling the true nature of organizational phenomena. First, a multilevel perspective has been widely adopted as scholars view organizational phenomena from a system perspectives, with lower level entities such as employees or teams nested in higher level entities such as units or organizations (cf., Hitt et al., 2007; Humphrey & LeBreton, 2019; Klein & Kozlowski, 2000; Mathieu & Chen, 2011). This multilevel theme is evident in different substantive domains such as emotions and affect (e.g., Ashkanasy et al., 2017), climate and culture (e.g., Schneider et al., 2017), collaboration (e.g., Bedwell et al., 2012), empowerment

This article was published Online First September 6, 2021.
John E. Mathieu https://orcid.org/0000-0002-6881-4022
This work was supported in part by grants from NASA grant #NNX09AK48G: *Cohesion Optimization in NASA Spaceflight Teams Experiencing Long-duration Assignment: Training and Indices for Optimizing Performance and Gaining Astronaut Synergy Under Stress,* Principle Investigator: Eduardo Salas, University of Central Florida; NASA grant # NNX11AR22G: *Composing and Developing Resilient, Adaptive, and Self-Sustaining Teams for Long Duration Space Exploration, Principle Investigator Scott Tannenbaum, The Group for Organizational Effectiveness;* and the Army Research Institute grant # W911NF-15-1-0014: *The Development and Construct Validation of Unobtrusive Dynamic Measures of Team Processes and Emergent States*, Principle Investigators Scott Tannenbaum and John Mathieu, The Group for Organizational Effectiveness, Inc. The opinions expressed herein, however, are strictly those of the authors and do not necessary represent the opinions of the sponsoring organizations.

The authors thank George Alliger, Chris Cerasoli, Lauren D'Innocenzo, Christina Lacerenza, Jamie Levy, Eduardo Salas, and Amanda Thayer for their efforts and contributions to this work. Some of these findings were presented in In J. E. Mathieu (Chair), Advanced Measurement of Team Dynamics. A symposium presented at the Annual Meeting of Society for Industrial and Organizational Psychology, Orlando, FL (April, 2017).

Correspondence concerning this article should be addressed to John E Mathieu, University of Connecticut Storrs, CT, United States. Email: John.Mathieu@uconn.edu

(e.g., Maynard et al., 2012), and leadership (Dionne et al., 2014). Second, there has been a growing appreciation that organizational phenomena are dynamic processes that change in meaningful ways over time (cf., Ancona et al., 2001; Mitchell & James, 2001; Shipp & Cole, 2015). Importantly, however, time is not a substantive variable per se, but rather, time serves as a surrogate variable for numerous temporal phenomena such as clock, cyclical, event, and life cycle-related processes (Ancona et al., 2001). Given that the passage of time will covary with qualitatively different temporal phenomena, it is important for scholars to develop and employ substantive measures of the nature(s) of the temporal dynamics that they hypothesize to be operating. And importantly, there may well be more than one temporal dynamic co-occurring over time. Regardless of the temporal theme(s) under consideration, organizational phenomena clearly need to be considered and modeled over time to understand their dynamic properties. Whereas both the multilevel and dynamic perspectives have generated abundant research applications, there have been very few studies that have integrated the two and modeled dynamic multilevel processes using data that are collected at a pace that is aligned with how quickly dynamic constructs are changing in the study context.

Despite the many calls for incorporating temporal perspectives into our research, there has been limited progress along these lines. For instance, scholars have expressed dissatisfaction about this limited progress that stems from, "a wide recognition of the need for new approaches that emphasize groups' nonlinear, complex, and dynamic nature, and the seeming inability of the field to adopt these approaches in its work" (Waller et al., 2016, p. 2). Shipp and Cole (2015) expressed similar sentiments exclaiming "organizational sciences have failed to heed such calls to adopt a temporal lens . . . [and] those who have studied time often do so independently of a specific research stream, and as a result, time has been applied haphazardly rather than systematically to organizational topics" (p. 238). Together, the multilevel and temporal perspectives suggest that organizational processes should be studied as complex multilevel processes that emerge and change over time. Unfortunately, our traditional methods of measurement are not well suited for such purposes. For instance, repeatedly surveying participants is subject to method-related factors such as response biases, survey fatigue, and testing effects. Observational methods provide rich insights concerning phenomena but are prone to biases of their own. In addition, the purpose of qualitative research is often to build contextualized theory, which efforts to scrub or sterilize the data for large-scale quantitative analysis would undermine (Pratt et al., 2020). Indeed, Kozlowski (2015) asserted "It is a simple fact that reliance on questionnaires as the dominant form of measurement for [collective constructs] team processes places significant limitations on efforts to capture process dynamics ... Thus, generally speaking, high frequency measurement that is low cost and unobtrusive is highly desirable" (pp. 278–279).

The overall purpose of our investigation is to demonstrate a method to use analyses of individuals' communications as construct valid unobtrusive measures of collective constructs occurring dynamically over time. For illustration purposes, we feature the development of such measures for a popular meso-level taxonomy of team processes proposed by Marks et al. (2001), yet the procedures that we document could be used to develop measures of virtually any level of organizational constructs, including macro-constructs such as human capital emergence (Ployhart & Moliterno, 2011) and organization-level psychological capital (McKenny et al., 2012). The Marks et al.'s (2001) framework is a dynamic model that features collective (team) constructs and is well established in the field, and as such serves as a fitting example to demonstrate the utility of our proposed computer-aided text analysis (CATA) approach.

We first briefly discuss the CATA approach and provide an overview of the Marks et al.'s (2001) team process framework. Second, we describe the development of a CATA protocol designed to yield indices of the Marks et al.'s dimensions based on team members' written or spoken (converted into written) language. Notably, we initially employ a content validity-based approach and supplement it with multimethod comparisons. Third, in Samples 1 and 2, we present empirical investigations based on video-taped recordings of three-member teams performing laboratory tasks. For both samples, participants' verbal communications were transcribed to text and then scored by trained subject matter experts (SMEs) and through CATA techniques. The convergent and discriminant validities of the scored transcripts were evaluated, along with other SMEs' ratings of the video-taped interactions, using multitrait–multimethod (MTMM) analyses. Fourth, in Sample 3, we present a third application comparing SME and CATA-scored transcripts of paramedic team members' communications during a series of high-fidelity mass casualty scenarios. Fifth, we further demonstrate the utility of CATA to model dynamic and multilevel constructs by examining team processes over time as related to team performance in a dynamic predictive model. We then illustrate how CATA can be scored at the individual level of analysis and integrated with wearable sensor data to generate scores that can be used to depict multidimensional dynamic team process networks over time. Finally, we conclude with implications, limitations, and directions for the future research, as well as guidelines for application.

Theoretical Background

Scholars have increasingly called attention to the value of focusing on behaviors and temporal behavioral contingencies (Lehmann-Willenbrock & Allen, 2018). However, work on these fronts has remained largely theoretical and conceptual (e.g., Waller et al., 2016), while empirical work to address and test these theories has been relatively sparse. Furthermore, most organizational behaviors are primarily measured by self-reports and it is difficult to judge the accuracy of such measures of behavior (Baumeister et al., 2007). If the intent is to study actual behavior through time, there is a need to conceptualize and measure markers that are as close to the intended behavior as possible (Lehmann-Willenbrock & Allen, 2018). Indeed, the benefit of studying actual behavioral markers is that the "obtained behavioral data are closer to the phenomena of interest, both conceptually and methodologically" (Lehmann-Willenbrock & Allen, 2018, p. 326). To demonstrate the utility of these kind of behavioral data, we focus on transcribed text, as words are powerful organizational tools that have been linked to a variety of outcomes ranging from consensus building to performance (Lockwood et al., 2019).

Historically, teams research has featured analysis of members' communications as a prime source of data (Mathieu et al., 2018). For instance, since Bales (1950) developed a protocol for observers to code team interactions in real time, numerous other systems have been developed (see Brauner et al., 2018). However, such protocols have typically been extremely time consuming and taxing efforts that involve recording and transcribing communications, establishing and refining coding schemes, and extensively training SME coders (Fischer et al., 2007). Naturally, such efforts provide deep understandings of team dynamics, but they are also arduous to produce and prone to intercoder unreliability (Seelandt, 2018).

In addition to the challenges associated with collecting and coding behavioral markers, researchers are faced with the challenge of how to align measures with the dynamics of the phenomena of interest. Luciano, Mathieu, et al. (2018) outlined two general strategies by which this can be accomplished. First, if one can anticipate a priori when shifts will occur from one stage to another, it may be possible to time measurements accordingly. For instance, observations, interviews, or surveys might be gathered as projects launch, transition through gateway reviews, or conclude their activities. However, that strategy poses several logistical challenges and transition points cannot always be anticipated and assessed. The second, and perhaps more viable strategy, may be to gather ongoing trace measures that can be used to index different constructs continuously over time.

Our goal with the current initiative is to provide a means of capturing the dynamics associated with collective phenomena, while balancing the benefits of extracting a rich understanding of members' interactions revealed through their words, with the desire to employ an efficient, reliable, valid, and scalable method to index such communications. We suggest that CATA offers a means to achieve our goal.

Computer-Aided Text Analysis

CATA is a general term used to describe a variety of different methods to analyze text with varying levels of fidelity (Short et al., 2018), ranging from simple counts of the relative use of singular versus plural pronouns (Steffens & Haslam, 2013), to the analysis of complex exchanges using natural language processing algorithms (Honnibal, 2016). Short et al. (2018) classify the techniques into two broad classes. One approach is akin to traditional content analysis and takes an inductive or grounded theory approach whereby patterns of word use are used to derive higher order meaning or themes that characterize a corpus of text. The second approach adopts a more deductive and content validity-based approach that analyzes the frequency of words in a corpus that corresponds to one or more predetermined word dictionaries that represent constructs of interest. In so doing, CATA explicitly focuses on members' communications and therefore provides a more proximal representation of their actual behaviors rather than a cognition or intention yielded by more traditional measurement approaches. For demonstration purposes, we sought a construct that would allow us to illustrate the value of CATA in a multilevel, dynamic way that is organizationally relevant and readily utilized.

Team Processes as an Exemplar Construct

Teams play an important role in organizations and often serve as the mechanism by which human capital resources are deployed to meet task demands (Ployhart & Moliterno, 2011). Work teams are "interdependent collections of individuals who share responsibility for specific outcomes for their organizations" (Sundstrom et al., 1990, p. 120). Research on work teams has flourished in the past few decades (Mathieu et al., 2017) and more recent frameworks feature dynamic and reciprocal relationships (cf., Ilgen et al., 2005; Mathieu et al., 2019), where team processes occupy a critical role whereby members work together to accomplish shared goals. Although teams are widely discussed as dynamic multilevel phenomena, due to overwhelming resource costs associated with traditional measurement techniques such as surveys and observations, they are often indexed at very limited number of time points.

The Marks et al.'s (2001) framework of team processes was selected as an exemplar framework as it is conceptualized to include complex temporal rhythms that may be difficult to truly capture with traditional approaches and it is well established in the teams literature. Marks et al. (2001) defined team process as:

> Members' interdependent acts that convert inputs to outcomes through cognitive, verbal, and behavioral activities directed toward organizing taskwork to achieve collective goals . . . Centrally, team process involves members interacting with other members and their task. They are the means by which members work interdependently to utilize various resources such as expertise, equipment, and money, to yield meaningful outcomes (e.g., product development, rate of work, team commitment, and satisfaction). (p. 357)

In addition, Marks et al. (2001) suggested that 10 lower-order processes (e.g., conflict management, strategy formulation, and coordination,) would map (10:3) to three higher-order team process constructs (i.e., transition, action, and interpersonal). In Table 1, we define the 10 lower-order team processes that map onto transition, action, and interpersonal processes. *Transition processes* occur prior to or between performance episodes when members reflect on previous experiences and plan for the future actions, including mission analysis, goal specification, and strategy formulation. Teams engage in *action processes* during performance episodes while working toward goal accomplishment, including monitoring progress toward goals, systems monitoring, team monitoring and backup behavior, and coordination. Finally, teams also engage in *interpersonal processes* to manage relationships between members at varying times, which could exist across episodes, and include conflict management, motivation, and confidence building, and affect management.

Notably, the Marks et al.'s (2001) framework has been cited over 3,622 times in Google Scholar as of June 2020. However, the studies that have sought to capture team processes have largely utilized survey measures, with few studies actually measuring multiple instances, and fewer still employing nonsurvey measures.[1] Stated differently, a key feature of the Marks et al.'s framework is that processes evolve and are linked over time, yet very few investigations have actually modeled them as dynamic processes. Indeed, Kozlowski (2015) submitted that "[t]eam processes are inherently dynamic phenomena theoretically, but they have largely been treated as static in research" (p. 270). This sentiment has been expressed by several others (Cronin et al., 2011; Marks et al., 2001; Mathieu & Luciano, 2019; McGrath, 1991) and remains true today. Marks et al.'s (2001) framework features performance episodes which "are distinguishable periods of time over which performance accrues and feedback is available (Marks et al., 2001, p. 359)." They may be considered in longer terms (e.g., quarterly sales goals and project phases) that may comprise shorter terms or subepisodes (e.g., weekly sales targets and

[1] Details regarding our review of literature citing and indexing one of more Marks et al.'s (2001) dimensions are available from the first author.

Table 1
Example Terms From Team Process Dictionaries

Dimensions	Definitions	Generic (N = 1,912)	Sample 1 (N = 301)	Sample 2 (N = 80)	Sample 3 (N = 4/16)[a]
Transition processes					
Mission analysis	Identification and evaluation of team tasks, challenges, environmental conditions, and resources available for performing team's work	Purpose N = 41	Protect N = 6	Log N = 4	
Goal specification	Activities centered on the identification and prioritization of team goals	Objective N = 23	Location N = 6		
Strategy formulation	Developing courses of actions, contingency plans, making adjustments to plans in light of changes, or expected changes in environment	Alternatives N = 81			
Higher order—transition		Procedure N = 345	Dock N = 7		*Who'll N = 5*
Action processes					
Monitoring progress	Paying attention to, interpreting, communicating information necessary for team to gauge progress toward its goals	Accomplish N = 97	Approach N = 79	Page N = 1	
System monitoring	Tracking team resources and factors in the team environment to ensure that the team has what it needs to accomplish its goals and objectives	Replenish N = 120	Charge N = 133		
Team monitoring and backup behavior	Members assisting others in the performance of their tasks (by providing feedback or coaching, or by assisting with the task itself)	Assist N = 78	Can-you N = 5	Received N = 3	*Back N = 1*
Coordination	The process of synchronizing or aligning members' actions	Co-operate N = 63	Align N = 62	Next N = 2	
Higher order—action		Do N = 398	Enemy N = 3	Bleeding N = 64	Treat/*Shoot* N = 4/*6*
Interpersonal processes					
Conflict management	The manner in which team members proactively and reactively deal with conflict	Mediate N = 64		Saw N = 1	
Motivation and confidence building	Activities that develop and maintain members' motivation and confidence while working toward team goals	Assure N = 110			
Affect management	Activities that foster emotional balance, togetherness, and effective coping with stressful demands and frustration	Enjoy N = 118		Down N = 2	
Higher order—interpersonal		Care N = 374		Drama N = 3	*Shot N = 4*

Note. Table presents positive words generated for each dictionary. Words in columns 3 through 6 are select example words for generic and supplemental dictionaries. The N below each example word reflects the numbers of words for that subdimension that exist in each dictionary.
[a] Italicized fonts in Sample 3 represent 16 "opt-out" words that have different meanings in that setting and were consequently excluded.

project tasks). Notably, teams likely perform multiple tasks that vary in duration such that they are concurrently engaging in different activities for one task than they are for other tasks (see Marks et al., 2001, Figure 1, p. 361). Although episode durations may vary, the important point is that they focus members' attention on the preparation and execution periods of activities.

Herein, we describe the process we used to develop word dictionaries for each of the dimensions, assess their content validity, and then, explore their convergent and discriminant validity versus SME hand-coded ratings of team communication transcripts, and SME ratings of video-recorded team interactions using Samples 1 and 2. In Sample 3, we then illustrate their use in a high-fidelity mass casualty incident exercise as well as their integration with wearable sensor badges to model dynamic networks of team interactions.

Dictionary Development

Short et al. (2010) advanced a method to develop construct valid CATA dictionaries. Given that we are beginning with a well-established conceptual framework, we adopted their deductive approach which entails as follows: (a) generating potential words for each dimension; (b) assessing the content validity of the words using SMEs; and (c) assessing the external validity of the resulting dictionaries. In our case, we conduct MTMM analyses of the dictionaries as compared to the SME hand-coded scoring of transcripts and a different set of SMEs' ratings of team processes based on video recordings and using Behaviorally Anchored Rating Scale (BARS). As detailed below, similar to the strategy advocated by Mathieu et al. (2020), as we seek to develop dictionaries for each of the 10 lower-order constructs articulated by Marks et al. (2001), we also identify methods to index the three higher-order constructs. In this manner, we develop methods to index team process at differing levels of fidelity or granularity for different research purposes.

We should mention two general themes before detailing our procedure. First, as noted by Short et al. (2018), many words may be viewed as indicators of different constructs, especially when the constructs are themselves highly related, such as the lower-order dimensions within each of Marks et al.'s three higher-order dimensions. For instance, the word *prepare* would clearly describe a team's transition processes but might also relate to mission analysis, goal specification, and/or strategy formulation and planning. In contrast, the word *plan* would more clearly belong to the last lower-order dimension (strategy formulation and planning). Hence, both during the word generation phase and later when reviewing word categorizations, words that clearly mapped to a single lower-order dimension were designated accordingly (Colquitt et al., 2019). In contrast, words that appeared to relate to multiple lower-order dimensions (e.g., conflict management and affect management) within the same higher-order dimension (i.e., interpersonal processes) were categorized as direct indicators of the higher-order construct. Other words that lapsed across different higher-order constructs (e.g., *teamwork*) were designated to an overall category.

A second guiding theme concerned the generality versus specificity of particular words. For instance, the word *energy* may describe members' motivational processes, a patient's or opponent's condition that should be considered during planning, or a vital resource (e.g., power consumption) that needs to be monitored during action processes. Therefore, we developed generic team process dictionaries designed to be widely applicable across team contexts with the expressed intent that they would be supplemented and customized for different team types or situations. The supplements feature idiosyncratic words and phrases unique to those team settings, and words that may connote something other than their typical meaning.[2]

Content Validity

In creating the CATA dictionaries for generic team processes, the second and third authors began with definitions for each of the 10 dimensions from the Marks et al.'s (2001) taxonomy along with existing measures (e.g., Mathieu et al., 2020) to clarify the construct domains. The authors then generated a list of words that describe positive team processes at the lowest level of specificity (i.e., the 10 lowest-order dimensions). Once these words were generated, the authors used a thesaurus (www.thesaurus.com) to seek out synonyms and antonyms of each of the initially generated words. These, in turn, were used to identify additional potential words, and the iterations varied for each word until the synonyms and antonyms were redundant. This process resulted in 2,518 words with some of the words repeated across lower-order dimensions. An additional 123 supplemental words were generated using frequency counts from pilot team transcripts from our three samples described below. Team process-related words that were deemed to be generic, rather than context specific, were subsequently added to the original word lists. After 416 repeated words were removed, a total of 2,225 words remained.

Next, seven different SMEs including the fourth and eighth authors were asked to categorize each word into any of the 10 lower-order dimensions that they thought it best corresponded to. If the word did not fit into only one lower-order dimension, they were allowed to place it into multiple ones. Following Anderson and Gerbing's (1991) procedures, we calculated coefficients of substantive validity (C_{sv}) for each of the generated words. C_{sv} values represent the percentage of time a given word was categorized as belonging to its intended category as compared to the next most frequently used category (see Anderson & Gerbing, 1991; Colquitt et al., 2019). We retained words that had a C_{sv} greater than .50 with their intended categories. Words that did not achieve the threshold for a lower-order category, but did meet it for a higher-order category, were in turn categorized into the corresponding higher-order category. Words with $C_{sv} \leq .50$ were revisited by the first two authors and one of the SMEs who assisted in the sorting task, who sought consensus concerning their placement. They agreed on the placement of 464 words and removed 120 words that were deemed as not fitting into any team process category. This process yielded 1,879 words that were further reviewed by the first two authors as a final check. They removed 79 words that they believed were ambiguous resulting in a final list of 1,800 words that we dubbed Version 1.0. Table 1 includes example words that are representative of the 10 lower-order

[2] When using the context-specific plugins, the context may alter the categorization of certain words. Consequently, we recategorized or removed words from the generic dictionary in certain instances for Samples 1 through 3. For Sample 1, 32 words were recategorized (e.g., Abandon, Enemy, and Fought). For Sample 2, 10 words were recategorized (e.g., Accident, Pain, and Sensitive). For Sample 3, 42 words were recategorized (e.g., Pain, Scream, and Terror).

dimensions, as well as the three higher-order dimensions of the Marks et al.'s (2001) team process taxonomy.

We next began using these words in three different applications and creating additional specific supplements for use in each environment (as detailed below). As our exemplar construct features an established framework, we first adopted a deductive approach, and then turned to an inductive supplement by exploring words existing in various prepublished dictionaries, which had potential for nomological overlap with our target categories (i.e., Hart et al., 2015; Pennebaker et al., 2015). Jointly with the unique applications and existing dictionaries, we generated an additional 1,504 words and categorized them accordingly using consensus among the third author and two previously unused SMEs. Consequently, Version 2.0 features 3,304 total words across 10 lower-order, 3 higher-order, and 1 overall dimension of the Marks et al.'s (2001) team process taxonomy. For the purposes of the current application, we set aside negative words as well as words referring to the overall team processes dimension, which resulted in a total of 1,912 context-generic terms.[3] These positive context-generic terms from the Version 2.0 dictionary (with supplements described below) were used for all three empirical examinations.

Empirical Investigations

Video-based observational methods enable raters to watch a sequence of activities, multiple times if necessary, and derive an overall impression of how the team executed different processes (Christianson, 2018). Unfortunately, they are also prone to rater halo effects, and performance cuing effects whereby observers may attribute team processes based on readily available performance information (McElroy & Downey, 1982). Observational ratings also require substantial SME time commitment and necessitate video recordings of team interactions, which may be intrusive and difficult to collect in many circumstances. SMEs' hand-coding of team transcripts also enables a deep understanding of team processes in context, and the ability to review previous passages and see communications in their full context. Because such coding is done on a line-by-line basis, it is less prone to halo and performance cuing effects and we consider it to be the "gold standard" by which to gauge the convergent validity of our CATA scoring (Reed et al., 2018). It is also the case that SME hand coding of transcripts is an extremely laborious, time consuming, and expensive method to index team processes. Accordingly, our primary research question is as follows:

> *To what extent do CATA-scored transcripts of team processes correlate with SMEs': (a) ratings based on video-recordings and (b) hand-coded transcripts?*

Sample 1

Participants

One hundred fifty undergraduate students from subject pools of two southeastern universities participated in the study (the University of Central Florida Institutional Review Board (IRB) Number: SBE-13-09231, Composing and Developing Resilient, Adaptive, and Self-Sustaining Teams for Long Duration Space Exploration). The sample was 52% female, had an average age of 19.39 years ($SD = 2.22$), and was 60% Caucasian, 17% Hispanic, 14% African American, and 9% "other." Participants were randomly assigned to 50 three-person teams, and to one of three positions within each team. These data represent a subsample of a larger investigation for which complete data for all three measurement protocols (CATA, hand-coding, and BARS) were available.

Experimental Task

Platform. The experimental platform was called Artemis, which is a PC-based simulated space exploration task and required collaboration and communication between team members. Specifically, Artemis is a multiplayer collaborative game where players occupy different roles and must work together to travel from one point to another, while potentially encountering obstacles or enemies. In our configuration, each team consisted of three highly interdependent members: (a) helm—primarily responsible for steering the ship and docking at space stations; (b) engineer—primarily responsible for allocating ship's energy, which is required for all system functioning; and (c) weapons—primarily responsible for managing defensive shields, and targeting and firing upon enemy ships. The primary coordination challenge for the team involved allocating and balancing ship energy use to sustain the helm and weapons needs.

Each participant was situated at a control center, which consisted of a desktop computer, monitor, keyboard, and mouse that controlled the simulation tasks and communication system. Participants wore microphone-equipped headsets and communicated through networked channels via a push-to-talk system that allowed them to speak with individual teammates or the entire team. All audio communication and participants' screens were recorded.

Missions. We scripted three 20-minute missions designed to test teams' processes in response to different environmental circumstances. These missions, therefore, constitute different self-contained episodes of activity (Marks et al., 2001) and serve as the focal units of analysis for our purposes. During the first mission, teams were tasked with traveling to, and successfully docking at, both an intermediate and a final base. They encountered two enemy ships along the way. Participants earned points for navigating and docking their ship, destroying enemy ships, and minimizing the damage that they incurred.

The second mission involved rescuing a distressed ship that needed help. Teams again had to navigate to and dock at the intermediate base, at which time they were informed that they need to find and escort an allied ship. This unexpected requirement served as stressor—an unexpected environmental challenge—that required problem solving and decision making among members. Along with escorting the allied ship, the team needed to destroy a total of six abandoned bases using a specific weapons system to successfully complete the mission. In short, the latter portion of this mission was a complex and daunting task designed to test their

[3] Short et al. (2018) also highlight an unresolved issue concerning the polarity of constructs. In other words, whether words should be generated to represent the content domain or its inverse, akin to negatively worded survey items. We actually sought to generate both positively and negatively worded item lists for each construct, and found that the negatively worded dictionaries were not negatively correlated with the positively worded dictionaries and otherwise behaved other than what we anticipated. We return to this point in the discussion.

transition and action processes while taxing their interpersonal processes.

For the third mission, the teams were again to navigate and dock at an intermediate base, after which they were to travel to and dock at a second base. However, exiting the first base, teams entered a nebula which undermined their navigation and defense capabilities and they had to traverse through a minefield. They encountered unexpected enemies while exiting the nebula prior to approaching their final base. In summary, the third mission presented complex demands that the teams were not prepared for and, again, put a premium on their team processes, but in a different manner than experienced in the second mission.[4]

Procedure

Each experimental session ran approximately two and a half hours including an initial training period and three episodes of game play (i.e., missions). Training modules were designed to provide participants with an overview of the game, the operations and requirements of their respective roles, how to communicate with other players, and how to read the game's map. Participants were also provided with a reminder information sheet for use during the simulation. At the end of training, a quiz was administered to ensure that all participants understood the simulation goals and how to perform their role before proceeding, and all participants passed.

Measures

SME BARS Ratings. SMEs watched the audio and video recordings and rated teams' processes using 10 BARS corresponding to the 10 lower-order Marks et al.'s (2001) dimensions. The BARS were rated using scales of one to five, with higher values representing more effective processes. Raters were provided with extensive frame-of-reference training to ensure they had a shared understanding of the construct(s). Using pilot data, raters discussed and clarified ratings until they had reached an acceptable level of interrater reliability. Each team-performance episode was randomly assigned to be rated by two of three SMEs. We tested and found no significant coder differences in mean ratings, and their average intercoder reliabilities were $r = .87$, $p < .01$; $r = .89$, $p < .01$; and $r = .50$, $p < .01$ for transition, action, and interpersonal processes, respectively. Therefore, we averaged the pairs of ratings, per episode, to index each team process. Notably, the 10 BARS were highly correlated within higher-order processes, so we averaged them to represent the higher-order constructs: (a) transition (3 items, $\alpha = .92$); (b) action (4 items, $\alpha = .94$); and (c) interpersonal (3 items, $\alpha = .90$).

SME Hand-Coded Transcripts. Participants' voice communications were recorded and transcribed into text documents. Following standard protocols (Reed et al., 2018), two trained graduate student SMEs reviewed the transcripts and separated them into stand-alone coherent statements. In other words, reading through the transcripts in context, the SMEs sorted out passages from a given member that contained multiple distinguishable thoughts in a single exchange into separate lines, and combined multiple utterances or statements that addressed or conveyed the same thought. This represents a classic syntax-based approach to segment the communications and focuses attention on the smallest discriminable portion of communication that observers can reliably classify (see Reed et al., 2018). Consequently, the resulting unit of analysis for communication coding became a single unique and meaningful thought or comment. The sequencing of communications was not altered in any way such that the resulting text revealed the ebb and flow of members' conversations. The SMEs then read the transcripts and coded each line of text into 1 of the 10 Marks et al.'s (2001) team process dimensions or not. "Ninety-two percent of the statements were coded in terms of team processes, whereas 8% were deemed to refer only to task information or were not codable." Coder calibration, interrater reliability, and scoring protocols are described below.

SME Calibration and Interrater Reliability. For pilot coding and SME calibration purposes, we used transcripts from 30 other performance episodes that contained 3,200 lines of text. First, the SMEs jointly reviewed a subsample of these pilot transcripts and discussed how various statements aligned with the Marks et al.'s 10 substantive lower-order team process dimensions and how those in turn mapped to the three higher-order process dimensions. For example, statements concerning *transition processes* included the following: (a) "Engineering, can you divert power from shields and maybe that gives us a little bit of energy?" (mission analysis); (b) "Does anyone know what we're supposed to do when we go through a nebula?" (goal specification); and (c) "Helm, let me know if you see any asteroids or anything, for the shields" (strategy formulation and planning). Example statements for *action processes* included the following: (a) "We're like halfway there, almost" (monitoring progress toward goals); (b) "Warp is now at 50%" (systems monitoring); (c) "I'm giving you additional power on the warp" (team monitoring and backup behavior); and (d) "Okay helm, what we're gonna do is we're gonna go all the way left until into A-1 and we're just gonna go straight down all the way to E-1" (coordination). Example statements for *interpersonal processes* included the following: (a) "That's my fault I'm sorry" (affect management); (b) "Oh god, we can do it. We can do it, we can do it. It's easy. We got it, we got it." (motivation and confidence building); and (c) "I've heard that before that's not helping though" (conflict management).

Using the 10 lower-order Marks et al. dimensions as categories, the two SMEs evidenced 79% exact agreement, which yields a kappa interrater reliability coefficient of $k = .77$, $p < .001$. Kappa represents the percentage of times SMEs classified statements as belonging to the same dimension minus the percentage of chance agreement (10%), divided by 1 minus the chance percentage. Generally, Kappa values .4–.6 are considered as *fair*, .6–.8 as *substantial*, and >.8 as *very good* (Landis & Koch, 1977). Most of the SMEs' disagreements herein, however, were for classifications that were clustered within the three higher-order dimensions. When using the three higher-order dimensions as categories, the two SMEs had 91% exact agreement and a $k = .87$, $p < .001$.

[4] For purposes of evaluating the psychometric properties of our measures, we consider the team episodes to be independent samples of team functioning that can be evaluated using multiple methods. For instance, we intentionally had the BARS coders watch episodes out of sequence and even mixed teams so as to minimize any carry over effects. Whereas the transcript coders saw the entire episode communications so that they could fully appreciate context within each episode, they too were assigned team transcripts in more of a random fashion to minimize temporal effects. Thus, we believe that treating each scored team episode as an independent observation of a team activity is reasonable for testing psychometric properties and MTMM relations.

Accordingly, each SME then coded 50% of the 150 team performance episodes and discussed any questionable issues with the other coder and reached consensus.

Given the overlap in word categorizations across Marks et al.'s lower-order dimensions within the same higher-order dimensions, and our desire to assess the convergent and discriminant validity of scores at the more general level of granularity, we summed the lower-order dimension scores to form transition, action, and interpersonal composites. Because count data are not presumed to yield parallel indicators of a given domain (i.e., tau-equivalent measures), we estimated the internal consistencies of these nonparallel (i.e., congeneric) measures (McDonald, 1999) using omega which were as follows: (a) transition (3 indicators) $\omega = .67$; (b) action (4 indicators) $\omega = .63$; and (c) interpersonal (3 indicators) $\omega = .68$.

CATA-Scored Transcripts. In addition to generating and validating the generic team process dictionaries, the first two authors produced a custom supplement of dictionaries that corresponded to the Artemis task. They first generated a list of relevant terms from the simulation documentation. Next, they conducted a frequency analysis on the terms from the pilot team transcripts. The list of words from these two approaches were then sorted into either the 10 lower-order dimensions, 3 second-order dimensions, or an overall team process dimension, resulting in 301 additional task-specific words that did not overlap with the generic list. The 150 team process episodes were analyzed using the CATA software Linquisitc Inquiry and Word Count (LIWC) 2015 (Pennebaker et al., 2015) with the combination of the generic- and task-specific team process dictionaries. Similar to other CATA software, LIWC counts the number of times words corresponding to specific dictionaries are mentioned in a given corpus. To align with the BARS and SME hand-coded scores, we aggregated the lower-order scores per higher-order dimension to form composites. Notably, these also included the scores for the words used as direct indicators of the second-order dimensions. The omega reliabilities for our unit weighted second-order CATA composites were as follows: (a) transition (4 indicators) $\omega = .76$; (b) action (5 indicators) $\omega = .69$; and (c) interpersonal (4 indicators) $\omega = .66$.

Sample 1 Results

We present descriptive statistics and correlations among the three second-order team process dimensions, each measured using the three different methods shown in Table 2. Of particular interest are the outlined validity diagonals where the same traits (i.e., process dimensions) are measured by different methods. Notably, the average dimensional correlations between BARS-rated and SME hand-coded transcripts were significant ($r = .25$, $p < .01$), as were those between the BARS and CATA-scored dimensions ($r = .32$, $p < .001$). Most encouraging, however, was the average correlation of $r = .76$, $p < .001$ between hand-coded and CATA-scored transcripts.

The validity diagonals yielded correlations supportive of convergent validity across methods as they were both significant and sufficiently large to encourage further examination (Campbell & Fiske, 1959). Although the MTMM approach developed by Campbell and Fiske (1959) can be informative, it has been criticized for (a) not accounting for the lack of independence of correlations

Table 2
Sample 1 Descriptive Statistics and Correlations

Dimensions	1	2	3	4	5	6	7	8	9	10	11	12
1. Transition_B	(.92)											
2. Action_B	.98***	(.94)								B-H: Mean $r = .25$, $p < .01$		
3. Interpersonal_B	.71***	.77**	(.90)									
4. Transition_H	.08	.04	-.03	(.67)						B-C: Mean $r = .32$, $p < .001$		
5. Action_H	.39***	.41***	.29***	.52***	(.63)							
6. Interpersonal_H	.30***	.32***	.26**	.52***	.72***	(.68)				H-C: Mean $r = .76$, $p < .001$		
7. Transition_C	.39***	.39***	.26**	.59***	.83***	.68***	(.76)					
8. Action_C	.32***	.34***	.23**	.63***	.90***	.67***	.89***	(.69)				
9. Interpersonal_C	.31***	.34***	.22**	.66***	.83***	.80***	.84***	.89***	(.66)			
10. % Women	-.10	.06	-.08	.15	.11	-.06	.02	.14	.02	-		
11. Average Age	-.21*	-.27**	-.13	.30**	-.02	.05	-.00	-.08	.01	.10	-	
12. Ethnicity Blau	-.18	-.08	-.23**	-.17	-.12	-.05	-.08	-.11	-.06	.24**	.39***	-
Means	2.95	3.32	3.34	13.67	60.45	18.30	36.32	85.76	46.76	19.45	50.97	.39
Standard Deviations	.87	.81	.62	13.42	38.81	16.75	27.82	57.77	36.29	1.72	30.26	.24

Note. $N = 150$ Team Performance Episodes. * $p < .05$. ** $p < .01$. *** $p < .001$.
B = SME BARS Ratings, H = SME Hand-Coded, C = CATA-Scored.
Rectangles highlight scale convergent validities with their averages provided in the three summaries on the right.
Numbers in parentheses represent reliability diagonal with Cronbach's alpha for BARS and omega for Hand-Coded and CATA-scored. See the online article for the color version of this table.

used in various comparisons; (b) lacking precise estimates of the amounts of trait- and method-related variance; and (c) not accounting for the distortion of correlations based on the measures' reliability (Widaman, 1985). As such, we performed a more rigorous test of the measurement properties using a series of nested confirmatory factor analyses (CFAs) of the MTMM correlations shown in Table 2 following Widaman's (1985) study. Note that Table 2 includes correlations with some team demographic composition variables for completeness that we do not consider further. We show a summary of our model comparisons in Table 3. We interpreted model fit using the standardized root mean square residual (SRMR) and the comparative fit index (CFI; Bentler, 1990), and include chi-square values that are useful for comparing the fit of nested models. Following Mathieu and Taylor's (2006) recommendations, we consider models with: CFI values <.90 and SRMR values >.10 as *deficient*, those with CFI ≥.90 to <.95 and SRMR >.08 to ≤.10 as *acceptable*, and those with CFI ≥.95 and SRMR ≤.08 as *excellent*.

Whereas traditionally MTMM applications are intended to test distinct traits and methods, in our case, we anticipated that both traits (i.e., team processes) and methods (i.e., SME BARS ratings, SME hand-coding, and CATA scoring) would be significantly correlated (Lance et al., 2002). Previous research has demonstrated that team processes are highly interrelated (Mathieu et al., 2020), and we are interested in how well the CATA scoring correlates with the SME hand-coded transcripts and with the video-based BARS ratings. Accordingly, we follow a model building protocol comparing nested CFA models to make those contrasts (see Cote & Buckley, 1987; Widaman, 1985).

First, we fit an uncorrelated-traits uncorrelated-methods (UTUM) null model [$\chi^2_{(20)} = 345.69$, $p < .001$; CFI = .716, SRMR = .335], which evidenced a deficient model fit but serves as a baseline comparison. Second, we fit an uncorrelated-traits, correlated-methods model [UTCM: $\chi^2_{(17)} = 99.48$, $p < .001$; CFI = .941; SRMR = .064], which evidenced an acceptable model fit and was significantly better than the UTUM model [$\Delta\chi^2_{(3)} = 246.21$, $p < .001$]. The UTCM versus the UTUM contrast provides evidence that there are significantly correlated methods of measurement. Third, we fit a correlated-traits, uncorrelated-methods model [CTUM: $\chi^2_{(17)} = 57.53$, $p < .001$; CFI = .971; SRMR = .077], which also exhibited an acceptable fit that was significantly better than the UTUM model [$\Delta\chi^2_{(3)} = 288.16$, $p < .001$]. This contrast is consistent with the expectation that the team process dimensions would be significantly correlated.

Finally, we fit a correlated-traits, correlated-methods model [CTCM: $\chi^2_{(14)} = 30.85$, $p < .001$; CFI = .988; SRMR = .084] which yielded an acceptable model fit that was significantly better than the UTCM model [$\Delta\chi^2_{(3)} = 68.63$, $p < .001$] and the CTUM model [$\Delta\chi^2_{(3)} = 26.68$, $p < .001$]. These findings suggest that the model including both correlated team process dimensions and correlated methods is warranted and fits well. The factor loadings from the CTCM model are shown in Figure 1. With the substantive relations simultaneously modeled, the BARS ratings correlated nonsignificantly with both the hand-coded transcripts ($\Phi = -.12$, ns) and the CATA-scored transcripts ($\Phi = -.11$, ns). Alternatively, the hand-coded and CATA-scored transcripts were highly correlated ($\Phi = .85$, $p < .001$). The substantive loadings of the CATA scores on the three team process dimensions were all substantial and significant as anticipated [i.e., transition: $\Gamma = .38$; action: $\Gamma = .50$; interpersonal: $\Gamma = .53$, $p < .05$]. Interestingly, the latent correlations among the process dimensions were large and highly significant (i.e., $\Phi_s = .82$, .88, and .89, $p < .001$) and comparable in magnitude to those reported by Mathieu et al. (2020) from surveys measures of team processes. To test the discriminant validity of the process dimensions, we fit a second CTCM model constraining the correlation between the latent transition and action processes (i.e., the highest of the three) to 1.0. This constrained model is nested in the CTCM model and evidenced a significantly worse [$\Delta\chi^2_{(1)} = 11.47$, $p < .001$] model fit [$\chi^2_{(15)} = 42.32$, $p < .001$; CFI = .980; SRMR = .040]. Constraining other pairings of latent process dimensions against the CTCM model yielded significantly worse model fits as well. Thus, although highly correlated suggesting the presence of methods variance, the latent team process dimensions do evidence significantly discriminant validity. Following Cote and Buckley's (1987) procedure to partition variance in the CTCM model, our analysis revealed that 49.72% of the total variance was attributable to team process dimensions, 46.83% to methods, and 3.45% to random error.

Summary

The findings from Sample 1 bode well for the use of CATA scoring of team processes. Using a deductive content validity-based approach, we developed generic and specific dictionaries for the 10 lower-order Marks et al.'s (2001) dimensions, along with 3 dictionaries corresponding to their higher-order dimensions. SMEs' BARS ratings of team processes exhibited during 150 episodes

Table 3

Alternative Sample 1 Multitrait Multimethod (MTMM) Model Tests

Model[a]	Traits	Methods	χ^2	DF	CFI	SRMR	$\Delta\chi^2$
1. UTUM	3 Uncorrelated	3 Uncorrelated	345.69***	20	.716	.335	—
2. UTCM	3 Uncorrelated	3 Correlated	99.48***	17	.941	.064	246.21***b
3. CTUM	3 Correlated	3 Uncorrelated	57.53***	17	.971	.077	288.16***b
4. CTCM	3 Correlated	3 Correlated	30.85***	14	.988	.084	68.63***c 26.68***d

Note. N = 150 team performance episodes; DF = Degrees of Freedom; CFI = Comparative Fit Index; SRMR = Standardized Root Mean Square Residual; UTUM = uncorrelated-traits, uncorrelated-methods; UTCM = uncorrelated-traits, correlated-methods model; CTUM = correlated-traits, uncorrelated-methods model; CTCM = correlated-traits, correlated-methods model.

[a] Fixed hand-coded transition and interpersonal, computer-aided text analysis (CATA) transition residual variances @ .00001.

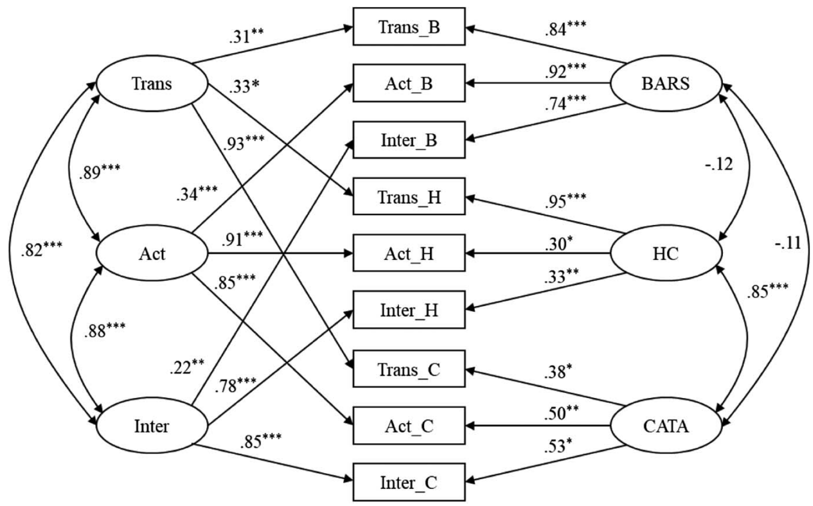
[b] Versus model 1, $\Delta DF = 3$.

[c] Versus model 2, $\Delta DF = 3$.

[d] Versus model 3, $\Delta DF = 3$.

** $p < .01$. *** $p < .001$.

Figure 1
Sample 1 Model 4 Confirmatory Factor Analysis (CFA) Results



Note. Trans = Transition Processes; Act = Action Processes; Inter = Interpersonal Processes. BARS (B) = Behaviorally anchored ratings scales; HC(H) = Hand-coded transcripts; CATA = Computer-aided text analysis.
N = 150 Team Performance Episodes; $\chi^2_{(14)} = 30.85^{***}$; comparative fit index (CFI) = .988; standardized root mean square residual (SRMR) = .084.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

evidenced significant correlations with CATA scoring of those interactions yet were nonsignificant when measurement factors were accounted for in the MTMM analyses. In contrast, the convergent validity of the CATA scoring with the time intensive and high fidelity "gold standard" SME hand-coding of team transcripts was significant, both in terms of correlations, and when modeled in the MTMM analyses (i.e., $\Phi = .85$, $p < .001$).

Although the findings chronicled above are encouraging, the observed correlations among the three higher-order Marks et al.'s (2001) dimensions were high and comparable to those found with survey measures (cf., Mathieu et al., 2020). These results may be attributable, at least in part, to the fact that all scorings were done from teams completing action phases of their task. In other words, teams conducted minimal transitional activities as they received mission briefings and objectives for each episode. Marks et al. (2001) and Luciano, Mathieu, et al. (2018) have stressed that it is important to assess team processes when they are actually occurring rather than concurrently at any given time. In other words, a better test of the construct validity of our measurement protocol alignment with Marks et al.'s theoretical framework would be to sample teams across transition and action periods and to discern whether our CATA measures are sensitive to the ebbs and flows of different processes at different times. As such, Sample 2 was designed to feature separate action and transition phases.

Sample 2

Participants

One hundred fifty-six undergraduate students from a large southeastern university were randomly assigned to 52 three-person teams and to one of three positions within each team (the University of Central Florida IRB Number: SBE-12-08298, Team Member Interactions and Team Performance). Note that other (nonoverlapping) data from this sample were featured in Bedwell's (2019) study. All participants received a cash stipend for their participation. The sample was 56% female, with 59% identifying as Caucasian, 20% as Latino/Hispanic, 16% as African American, and 5% distributed over other categories. Ages ranged from 18 to 57 years, with most participants (66%) ranging between 18 and 21.

Experimental Task

Teams engaged in a computer-based simulation set in an emergency room waiting area, filmed from a first-person view. Actors portrayed the role of doctors, other volunteers, and patients who were scripted to act in a particular manner. Participants "interacted" with the actors, simulating a real conversation even though it was recorded video—essentially, the video would pause after an actor posed a question to the participant and would only begin after the participant finished speaking an answer (see Smith-Jentsch, 2007 for more details). There were three participant roles: (a) Waiting Room Staffer; (b) Records Staffer; and (c) Claims Staffer. The Waiting Room Staffer interacted directly with the simulation, answering patient/staff questions, and responding to voicemails. The Records Staffer maintained an employee tracking form and a patient log form. The Claims Staffer completed an insurance claim form for patients, a complaint form (for formal complaints made against hospital employees), and received patient details from the admittance department (aka, experimenter) via chat. Performance was based on the degree to which participant teams correctly triaged all patients (i.e., rank ordered them as to who should be seen first, second, etc.). Given that each position held unique information about patients, successful team performance required effective information sharing among team members.

Procedure

Each experimental session ran approximately two and a half hours including an initial training period and a simulated scenario with transition and action phases. Participants were shown a training video (i.e., a voice-enhanced PowerPoint) that described the simulation, their roles, and associated tasks. Using a provided worksheet, teams engaged in a 15-min planning (i.e., transition) period, after which they completed an action phase. Forty-two of our sample teams had complete information available for both phases, yielding 84 episodes. An additional 10 teams completed a second round of the simulation which yielded another 40 episodes, and therefore, 124 episodes in total for analysis purposes.

Measures

SME BARS Ratings. Two SMEs discussed the coding strategy and then both independently rated 10% of the video-recorded experimental sessions. Based on this sample, we tested and found no significant coder differences in mean ratings, and their average intercoder reliabilities were $r = .92$, $p < .001$; $r = .67$, $p < .01$; and $r = .75$, $p < .001$ for transition, action, and interpersonal processes, respectively. Therefore, we averaged the pairs of ratings, per episode, to index each team process, and each SME rated 50% of the remaining episodes, and discussed any questionable instances. The 10 BARS ratings represented three higher-order processes: (a) transition ($\alpha = .83$); (b) action ($\alpha = .78$); and (c) interpersonal ($\alpha = .49$). Notably, the interpersonal process BARS showed very little variance attenuating their internal consistency estimate. The BARS higher-order dimension descriptive statistics and intercorrelations are shown in Table 4.

SME Hand-Coded Transcripts. As with Sample 1, two SMEs first reviewed the transcripts and separated them in terms of stand-alone coherent statements for analysis. For pilot coding and SME calibration purposes, we used transcripts from six performance episodes that contained 1,251 lines of text. The SMEs jointly reviewed a subsample of these pilot transcripts and discussed how various statements aligned with the Marks et al.'s 10 substantive lower-order team process dimensions and how those in turn mapped to the three higher-order process dimensions. For example, statements concerning *transition processes* included the following: (a) "Ok, so who wants to be the backup for the claims volunteer or staffer?" (mission analysis); (b) "Let's do, um, file formal complaints. And make insurance claim or fill insurance claim forms." (Goal specification); and (c) "How are we going to do that?" (strategy formulation and planning). Example statements for *action processes* included the following: (a) "Yeah, he did one." (monitoring progress toward goals); (b) "Did you get this list of patients.? (systems monitoring); (c) "Want me to take on the coda work sheet?" (team monitoring and backup behavior); and (d) "You are filling out the date of birth and stuff like that?" (coordination). And, example statements for *interpersonal processes* included the following: (a) "Aw you're crying." (affect management); (b) "Let's do this!" (motivation and confidence building); and (c) "Well yeah we have to talk about it as a team." (conflict management).

The SMEs then independently rated 475 pilot statements. Using the 10 lower-order Marks et al.'s dimensions as categories, the two

Table 4
Sample 2 Descriptive Statistics and Correlations

Dimensions	1	2	3	4	5	6	7	8	9	10	11	12
1. Transition_B	(.83)											
2. Action_B	–.01	(.78)										
3. Interpersonal_B	.22*	.17	(.49)									
4. Transition_H	.60***	–.27**	.01	(.65)								
5. Action_H	–.57**	.32***	–.00	–.84***	(.60)							
6. Interpersonal_H	.19*	-.01	.28***	.16	–.02	(.60)						
7. Transition_C	.18*	.27**	.18*	.14	.11	.25**	(.80)					
8. Action_C	–.07	.35***	.25**	–.17	.41***	.22*	.80***	(.69)				
9. Interpersonal_C	.11	.37***	.36**	–.01	.24**	.42***	.79***	.81***	(.75)			
10. % Women	–.11	–.03	.04	.05	.15	.08	.10	.19*	.24**	–		
11. Average Age	–.03	.07	–.17	.06	–.03	–.08	–.20*	–.21*	–.23*	–.05	–	
12. Ethnicity Blau	–.03	.18*	.22*	.07	.12	.27**	.14	.12	.22*	.20*	.13	–
Means	3.45	3.60	3.44	33.26	34.28	13.27	44.93	85.19	54.12	.55	21.56	.44
Standard Deviations	.53	.67	.47	46.53	49.88	20.72	29.89	50.25	37.30	.29	2.61	.20

B-H: Mean $r = .40$, $p < .001$

B-C: Mean $r = .30$, $p < .01$

H-C: Mean $r = .32$, $p < .001$

Note. $N = 124$ Team Performance Episodes. * $p < .05$. ** $p < .01$. *** $p < .001$.
B = SME BARS Ratings, H = SME Hand-Coded, C = CATA-Scored.
Rectangles highlight scale convergent validities with their averages provided in the three summaries on the right.
Numbers in parentheses represent reliability diagonal with Cronbach's alpha for BARS and omega for Hand-Coded and CATA-scored.
See the online article for the color version of this table.

SMEs evidenced 49% exact agreement, which yields a kappa interrater reliability coefficient of $k = .43$, $p < .001$. Most of the disagreements, however, were for classifications that were clustered within the three higher-order dimensions. Using the three higher-order dimensions as categories, the two SMEs had 89% exact agreement and a $k = .84$, $p < .001$. Of the total of 16,064 lines of text, 92% were coded in terms of team processes, whereas 8% referred to task information or were not codable.

Accordingly, each SME then coded 50% of the 14,779 lines of text from the 124 team performance episodes and discussed any questionable ones to reach consensus. We aggregated the lower-order scores per dimension and task phase, and we calculated the omega reliabilities as transition (3 indicators) $\omega = .65$, action (4 indicators) $\omega = .60$, and interpersonal (3 indicators) $\omega = .60$.

CATA-Scored Transcripts. We used Version 2 of the generic CATA dictionaries along with a custom supplement of words that corresponded to the task in Sample 2 generated by the first two authors. They began by generating a list of relevant terms from the documentation on the waiting room simulation. Next, they conducted a frequency analysis on terms expressed in the pilot transcripts. The list of words from these two approaches were then sorted, using consensus, into either the 10 lower-order dimensions, the 3 second-order dimensions, or the overall team process dimension. This resulted in an additional 80 task-specific words that did not overlap with the generic list. Then, using the combined generic and customized dictionaries, the 124 team process episodes were analyzed using the LIWC (Pennebaker et al., 2015) CATA software. The omega reliabilities for our unit weighted second-order CATA composites were as follows: (a) transition (4 indicators) $\omega = .80$; (b) action (5 indicators) $\omega = .69$; and (c) interpersonal (4 indicators) $\omega = .75$.

Sample 2 Results

Following the same analytic procedure as Sample 1, Table 4 shows descriptive statistics and correlations among Sample 2 variables. Validity diagonals reveal that the average correlations for the corresponding BARS-rated–SME hand-coded transcripts ($r = .40$, $p < .001$), the BARS–CATA-scored dimensions ($r = .30$, $p < .01$), the hand-coded-CATA-scored transcripts averaged ($r = .32$, $p < .001$) were all significant. The zero order correlations also suggested the likely presence of methods effects, particularly among the SME hand-coded and CATA-scored transcripts.

Moving to the MTMM–CFA analyses shown in Table 5, the UTUM null model [$\chi^2_{(19)} = 135.51$, $p < .001$; CFI = .819; SRMR = .165] evidenced a deficient model fit. Both the UTCM model [$\chi^2_{(16)} = 80.12$, $p < .001$; CFI = .900; SRMR = .120] and the CTUM model [$\chi^2_{(16)} = 66.62$, $p < .001$; CFI = .921; SRMR = .125] exhibited significantly better fit as compared to the UTUM model [$\Delta\chi^2_{(3)} = 55.39$, $p < .001$ and $\Delta\chi^2_{(3)} = 68.89$, $p < .001$, respectively], but were deficient on the basis of their SRMR indices. These contrasts provide evidence that the methods of measurement were significantly correlated with each other, as were the team process dimensions with each other, respectively. Finally, the CTCM model exhibited an excellent model fit [$\chi^2_{(13)} = 18.95$, ns; CFI = .991; SRMR = .048], which was significantly better than the UTCM model [$\Delta\chi^2_{(3)} = 61.17$, $p < .001$] and the CTUM model [$\Delta\chi^2_{(3)} = 47.67$, $p < .001$]. These findings suggest that the model including both correlated team process dimensions and correlated methods is warranted and fits well. The factor loadings from the CTCM model for Sample 2 are shown in Figure 2. With the substantive relations simultaneously modeled, the latent correlations between the BARS method and hand-coded transcripts were not significant ($\Phi = .16$, ns), whereas both the BARS and the hand-coded methods correlated significantly with the CATA-scored transcripts ($\Phi_s = .61$ and .74, $p < .001$, respectively). The substantive loadings of the CATA-scored team process dimensions were all substantial and significant as anticipated [i.e., transition: $\Gamma = .91$; action: $\Gamma = .90$; interpersonal: $\Gamma = .89$, $p < .001$]. The CTCM model revealed that 40.19% of the total variance was attributable to team process dimensions, 42.20% to methods, and 17.62% to random error.

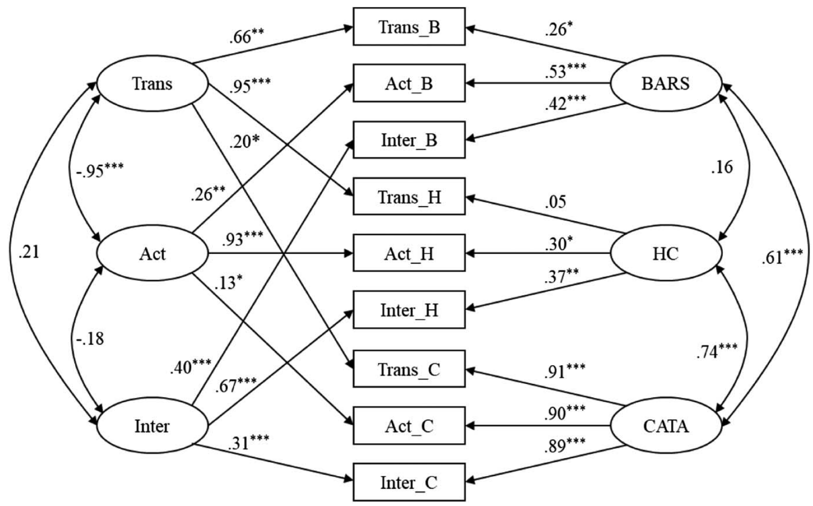
Interestingly, the latent correlations among the process dimensions were nonsignificant between the interpersonal processes and both transition ($\Phi = .21$, ns) and action processes ($\Phi = -.18$, ns), but was quite large, negative, and significant between the transition and action processes ($\Phi = -.95$, $p < .001$). Indeed, we fit a second CTCM method constraining the correlation between the latent transition and action processes to 1.0, which did *not* differ significantly [$\Delta\chi^2_{(1)} = .72$, ns] from the original CTCM model and evidenced excellent fit [$\chi^2_{(14)} = 19.67$, ns; CFI = .991; SRMR = .043]. Most likely this finding is attributable to the fact that teams were performing either dedicated planning or action episodes, which accounts for the exceptionally high negative

Table 5
Alternative Sample 2 Multitrait Multimethod (MTMM) Model Tests

Model[a]	Traits	Methods	χ^2	DF	CFI	SRMR	$\Delta\chi^2$
1. UTUM	3 Uncorrelated	3 Uncorrelated	135.51***	19	.819	.165	—
2. UTCM	3 Uncorrelated	3 Correlated	80.12***	16	.900	.120	55.39***b
3. CTUM	3 Correlated	3 Uncorrelated	66.62***	16	.921	.125	68.89***b
4. CTCM	3 Correlated	3 Correlated	18.95	13	.991	.048	61.17***c
							47.67***d

Note. N = 124 team performance episodes. DF = Degrees of Freedom; CFI = Comparative Fit Index; SRMR = Standardized Root Mean Square Residual; UTUM = uncorrelated-traits, uncorrelated-methods; UTCM = uncorrelated-traits, correlated-methods model; CTUM = correlated-traits, uncorrelated-methods model; CTCM = correlated-traits, correlated-methods model.
[a] Fixed hand-coded action residual variances @ .00001.
[b] Versus model 1, $\Delta DF = 3$.
[c] Versus model 2, $\Delta DF = 3$.
[d] Versus model 3, $\Delta DF = 3$.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Figure 2
Sample 2 Model 4 Confirmatory Factor Analysis (CFA) Results



Note. Trans = Transition processes, Act =Action processes, Inter = Interpersonal processes. BARS (B) = Behaviorally anchored ratings scales; HC(H) = Hand-coded transcripts; CATA = Computer-aided text analysis.
N = 124 Team performance episodes; $\chi^2_{(13)} = 18.95$, *ns*; comparative fit index (CFI) = .991; standardized root mean square residual (SRMR) = .048.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

correlation. In contrast, both transition and action processes evidenced significant discriminant validity from interpersonal processes.

Summary

As with Sample 1, SMEs' BARSs ratings and hand-coding of team processes exhibited during 124 episodes evidenced significant convergent validity with CATA scoring of those interactions. Again, in the context of the MTMM analyses, the convergent validity of the CATA SMEs' hand-coding of team transcripts was high $\Phi = .74$, $p < .001$. Interestingly, the correlations among the team processes were much different than those observed in Sample 1. Interpersonal processes did not correlate significantly with transition or action processes, whereas the latter two evidenced a substantial *negative* correlation. Given that we sampled teams in dedicated transition (i.e., planning) or action (i.e., task execution) phases, these findings are actually consistent with Marks et al.'s (2001) original conceptualization of episodic processes—that teams engage in different processes over time associated with task requirements.

Whereas the findings from Samples 1 and 2 are supportive of the use of CATA scoring of team communications, they were derived from small teams of unfamiliar undergraduates performing short-term laboratory tasks. Clearly, it is important to test the applicability of these measurements with teams performing dynamic real-world tasks, whose members know one another. Accordingly, Sample 3 was designed to provide such a test.

Sample 3

Sample 3 featured 13 paramedic trainees (hereafter paramedics) from an emergency medical response academy who participated in six live-actor simulations (referred to as scenarios) that each lasted between 20.75 and 30.00 ($M = 25.00$, $SD = 4.05$) minutes, over the course of a single day exercise. The data presented in this article (mass-casualty incident [MCI] exercise code 1603) were part of a broader data collection effort that involved several day-long training exercises, conducted over several years, with different participants completing different exercises (Arizona State University IRB #: 00004290, Exploring Multiteam Systems Dynamics in Emergency Medical Response to Mass Casualty Incidents). In our analysis, we first examine our primary research question: *To what extent do CATA-scored transcripts of team processes correlate with SME ratings* (here, hand-coded transcripts). We then expand our investigation to further illuminate the utility of CATA for exploring dynamic and multilevel phenomenon. First, we examine *team processes over time* by correlating episodic processes with their earlier counterparts from the previous episode (i.e., Time-1) for each of our samples. Second, Sample 3 provides an opportunity to explore the *team process–performance relationship over time* by associating temporal CATA measures of team processes with a dynamic criterion in a longitudinal predictive design. Finally, we also score CATA measures at the individual level of analysis and leverage wearable sensor data to illustrate how these measurement techniques can be used to generate time-based, dynamic, social *network representations of team processes* (Crawford & LePine, 2013).

Participants

The paramedics had been in class together one day per week for nearly 12 months. Their mean age was 26.6 ($SD = 6.8$) years and 46% were female. They reported their ethnicities as White or Caucasian (61.5%), Hispanic or Latino (23.1%), or other (15.4%). Their education levels ranged from high school (23.1%), to associate (46.2%), or bachelor (30.8%) degrees. All participants donned sociometric badges (Kim et al., 2012) that were used to track their proximity to one another, and personal voice-recording devices from which we generated individual time-stamped transcriptions of what each person said throughout the exercises. This approach allowed us to examine individuals' communication both as a team and as networks of individuals.

Task

The training exercise served as a capstone simulation for their paramedic training program and featured the characteristics of MCIs. MCIs are generally defined as "any incident in which emergency medical service resources, such as personnel and equipment, are overwhelmed by the number and severity of casualties" (Mistovich et al., 2013, p. 302). The MCI exercise included six scenarios throughout the day, three simulated multiple-motor vehicle incidents in the morning, and three simulated mass shooting incidents in the afternoon. The exercise occurred in the area of approximately 7,000 square feet that included several different rooms in the academy's building as well as part of an adjacent parking lot. Emergency Medical Technician trainees at the academy played the roles of patients ($N = 15-16$, per scenario) and "presented" symptoms associated with their respective scripted injuries, which were further represented by a moulage artist. Additional actors portraying different roles across scenarios (e.g., police officers) were included with special effects (e.g., noise), to add realism.

The ultimate goal of each MCI is to provide urgent medical care before transporting patients to their next medical care facility. The paramedics were deployed in three four-member teams, a formal leader and three members per team, and reported to a formal system leader (i.e., incident commander, IC). The three teams per scenario were as follows: (a) *Triage*—where the focus was to find all patients, accurately identify their medical conditions, and prioritize them for treatment; (b) *Treatment*—which established an initial staging area where patients were reassessed, received immediate care to stabilize them, and were sequenced for transport; and (c) *Transport*—where patients' needs and resource availability (e.g., ambulances) were coordinated and patients were removed from the scene to one or more hospitals. In effect, these teams worked as a small multiteam system (MTS), which "are tightly coupled networks of teams that pursue at least one shared superordinate goal in addition to their component team goals" (Luciano, DeChurch, & Mathieu, 2018, p. 1066). Given the sequential nature of operations, some members of triage and treatment later joined subsequent teams as the last patients exited from their areas of operations. Because of the uncertainty and dynamic changes in environment throughout the scenarios, paramedics were required to actively revise their strategy, coordinate their actions, and maintain morale.

Measures

Team Performance. Each scenario had 15 or 16 patients who presented symptoms that ranged in severity from: (a) minor injuries; (b) moderate injuries; (c) severe injuries; and (d) deceased. Protocol dictates that severely injured (but still savable) patients should have the highest processing priority, followed by those with moderate injuries, then minor injuries, and finally the deceased (City of Phoenix, 2000). Notably, moving patients in this sequence can be directly linked to saving lives and limbs (Pollak et al., 2012).

In the exercise, patients' level of injury severity was distributed throughout the incident region. This created team challenges for sequencing patients appropriately as it required participants to work together to assess and track the status of all patients in order to determine which patient should be addressed next (e.g., the severely injured patient in the black sedan should be moved to Treatment before the patients with minor injuries in the white truck). Consequently, as soon as patients arrived in a team's area of operation, we could calculate the accuracy of sequencing in each team's area at any given time. For example, if there were four patients present in the Treatment area and one severely injured patient was not being treated in the correct order (e.g., after a patient with minor injuries), the performance score would be .75 (three correctly sequenced/four total) at that time. We used a ratio of correctly sequenced patients divided by total patients, as the number of patients in each team's area varied over time (from 1 to 16). Patients were present in each team's area between 9.50 and 23 minutes per scenario ($M = 14.99$, $SD = 3.46$). This yielded 1,078 15-second segments, where performance scores ranged from 0 (*completely incorrect*) to 1.0 (*perfectly sequenced*) ($M = .71$, $SD = .34$).

SME Hand-Coded Transcripts. The 13 participants' voice recordings were converted into individual time-stamped transcriptions. We then created team interaction transcripts by combining the transcripts of members of the same team at any given point in time. Seven other SMEs reviewed the transcripts and separated them into stand-alone coherent statements ($N = 5,978$ lines of text). Of these, 98.8% of the statements were coded in terms of team processes, whereas 1.2% conveyed only task information or were not codable. For pilot coding and calibration purposes, we used transcripts containing 1,818 lines.

First, the SMEs jointly reviewed a subsample of those pilot transcripts and discussed how various statements aligned with the Marks et al.'s (2001) 10 lower-order team process dimensions and how those in turn mapped to the three higher-order process dimensions. Once the coding protocol was set, pairs of the SMEs independently coded each of the 1,818 lines of text into one of the 10 lower-order team process dimensions. Using these 10 dimensions as categories, the SMEs evidenced 84.5% exact agreement, which yields a kappa interrater reliability coefficient of $k = .83$, $p < .001$. Using the three higher-order dimensions as categories, the SMEs had 91% exact agreement and a $k = .86$, $p < .001$. Accordingly, SMEs then each coded 40%–50% of the 2,395 lines of text and discussed any questionable ones with the other coder and reached consensus.

CATA-Scored Transcripts. We employed a three-step approach to develop the MCI CATA supplement. First, using transcripts from a different MCI exercise, we identified 923 frequently used words. Second, we compared those words to our Generic Dictionaries (Version 2) and found only four that were deemed as unique to the MCI setting (i.e., treat, treating, triaging, and transporting—all higher-order action process words). Third, we reviewed words in our generic dictionaries that connote different meanings in an MCI context and excluded 16 that had different meanings in this context. The transcriptions were analyzed using LIWC CATA software (Pennebaker et al., 2015) and our generic Version 2.0 dictionaries as supplemented with the MCI-specific module.

Sample 3 scoring of our hand-coded and CATA-scored transcripts differed from Samples 1 and 2 in two important ways. First, we sought to discern more focused or fine-grained subepisodes of activities. Marks et al.'s (2001) temporal framework submitted that "Episodes' durations stem largely from the nature of the tasks that teams perform and the technology that they employ, and from the manner in which members choose to complete work. Episodes are most easily identified by goals and goal accomplishment periods" (p. 359). Accordingly, in Samples 1 and 2, we operationalized episodes as associated with distinguishable experimental tasks. Yet, in an MCI context, team activities were focused on finding, diagnosing, treating, and transporting patients—many of which were being handled simultaneously. Marks et al. (2001) submitted that episodes can vary in length and consistency, and they may often be segmented into subepisodes of more focused work that contributes to the larger effort. Moreover, "teams have to multitask in order to manage several performance episodes simultaneously (McGrath, 1991). Consequently, they often work in multiple performance episodes at a given point in time, each with its constituent subgoals and episodes and with its associated rhythms and sequence" (Marks et al., 2001, p. 360).

Given the theoretical prescriptions of Marks et al. (2001), it is potentially valuable to index team processes in a more fine-grained time sensitive manner. Luciano, Mathieu, et al. (2018) argued that "determining the appropriate temporal unitization requires exploring what the smallest meaningful samples of behavior, cognition, or affect are necessary to yield valid snapshots of the construct, which guides decisions as to how to aggregate the data over time" (pp. 608–609). Based on a qualitative analysis of the pace of MCI activities, we decided to index team processes per minute ($N = 476$ episodes).

Accordingly, we calculated hand- and CATA-scoring of participants' communications per minute to score their 10 lower-order and 3 higher-order Marks et al.'s dimensions. The *team-level* omega reliabilities for the hand-coded scores across the 1-min episodes were as follows: (a) transition (3 indicators) $\omega = .56$, action (4 indicators) $\omega = .52$, and interpersonal (3 indicators) $\omega = .53$, whereas the corresponding omegas for the CATA scores were as follows: (a) transition (4 indicators) $\omega = .77$, action (5 indicators) $\omega = .48$, and interpersonal (4 indicators) $\omega = .77$. Notably, these reliabilities' values are highly restricted as compared to Samples 1 and 2 as they were derived from limited score ranges given the short durations over which they are accumulated.

Sample 3 Results: Convergent and Discriminant Validity

Table 6 shows descriptive statistics and correlations among the Sample 3 variables. SME hand-coded scores correlated significantly with the CATA-scored transcripts for all three process dimensions: (a) transition: $r = .53$, $p < .001$; (b) action: $r = .57$, $p < .001$; and (c) interpersonal: $r = .38$, $p < .001$, averaging $r = .49$, $p < .001$.

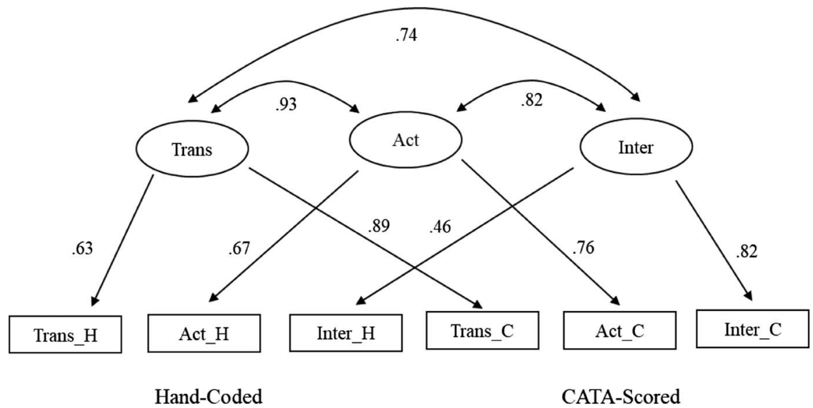
Because there are not sufficient degrees of freedom to conduct MTMM analyses with only three substantive and two method dimensions, we used the corresponding hand-coded and CATA-scored indices per Marks et al.'s dimension and estimated a three-factor model. This model exhibited excellent fit indices [$\chi^2_{(6)} = 17.06$, $p < .01$; CFI = .989; SRMR = .021]. As shown in Figure 3, all indicators loaded significantly on their intended latent variables, ranging from $\Gamma = .46$, $p < .001$ to $\Gamma = .89$, $p < .001$, averaging $\Gamma = .70$, $p < .001$. The correlations among the latent variables were all significant and large: (a) transition–action: $\Phi = .93$, $p < .001$; (b) transition–interpersonal: $\Phi = .74$, $p < .001$; and (c) action–interpersonal: $\Phi = .82$, $p < .001$, averaging $\Phi = .83$, $p < .001$. We then fit a two-factor model constraining the correlation between the latent transition and action processes to 1.0, which fit well [$\chi^2_{(7)} = 21.51$, $p < .01$; CFI = .985; SRMR = .023], but was significantly worse [$\Delta\chi^2_{(1)} = 4.45$, $p < .05$] than the

Table 6
Sample 3 Team Level Descriptive Statistics and Correlations

Dimensions	1	2	3	4	5	6
1. Transition_H	(.56)					
2. Action_H	.37***	(.52)		H-C: r = .49, p < .001		
3. Interpersonal_H	.23***	.26***	(.53)			
4. Transition_C	$\boxed{.53^{***}}$	.54***	.22***	(.77)		
5. Action_C	.42***	$\boxed{.57^{***}}$	.29***	.65***	(.48)	
6. Interpersonal_C	.41***	.52***	$\boxed{.38^{***}}$	.46***	.46***	(.77)
Team-Level						
Means	2.06	9.02	.59	4.10	5.05	2.91
Standard Deviations	2.31	7.14	1.22	4.16	4.65	2.74

Note. N = 476 Team Minute Episodes. * $p < .05$. ** $p < .01$. *** $p < .001$.
H = SME Hand-Coded, C = CATA-Scored.
Rectangles highlight scale convergent validities with their average provided in the summary on the right.
Numbers in parentheses represent reliability diagonal with omega for Hand-Coded and CATA-scored.

Figure 3
Sample 3 Team-Level Three-Factor Model Results



Note. Trans = Transition processes; Act =Action processes; Inter = Interpersonal processes.
H = Hand-coded transcripts; C = Computer-aided text analysis.
Standardized Factor Loadings shown, $p < .001$.
Values from $N = 476$ team minute episodes.

three-factor model. Thus, although highly correlated, these findings provide evidence of the discriminant validity among the process dimensions.

Team Processes Over Time. The MTMM analyses that we conducted with Samples 1–3 compiled multiple instances where indices of team processes were collected concurrently during the same periods. Given that there were no dedicated planning and action phases in either Sample 1 or 3, the correlations among their three second-order team process dimensions were positive and quite high. In contrast, transition and action processes exhibited strong negative relations in Sample 2. On point, LePine et al. (2008) submitted "we encourage future research to employ time-based research designs whereby measures of different team processes can be aligned with when they are anticipated to occur" (p. 297). To shed light on the temporal nature of team processes, we correlated episodic processes with their earlier counterparts from the previous episode (i.e., Time-1) for each of our samples as shown in Table 7. In other words, those correlations depict team process relationships across adjacent episodes.

The three processes all intercorrelated with one another with rs = .52–.69, $p < .001$ using Sample 1 data. Recall that these teams were performing a series of action phases where they were provided with information, goals, and strategies. Essentially these correlations represent the consistency of team processes exhibited over a series of action phases. In contrast, Sample 2 correlations showed less consistency over time, which follows from the fact that they performed qualitatively different functions over time—a dedicated planning session (episode), followed by an action session (episode). As shown, prior transition processes correlated positively with subsequent action processes ($r = .49$, $p < .01$), which is in stark contrast to their massive negative concurrent correlation ($\Phi = -95$, $p < .001$) seen in the MTMM analysis. These findings are consistent with the Marks et al.'s (2001) thesis that processes are linked over time in an episodic fashion.

Sample 3 data evidenced a different pattern over time. Whereas all three processes exhibited significant consistency over time, prior transition processes related positively to subsequent action processes ($r = .44$, $p < .001$), and prior action processes related positively to subsequent transition processes ($r = .42$, $p < .001$). Consistent with Marks et al.'s (2001) thesis that teams simultaneously orchestrate multiple tasks that are not necessarily synchronized (i.e., treating patients in Sample 3), these findings suggest that teams are engaging in an ongoing series of different task episodes that do not necessarily align temporally. Interpersonal processes correlated the least with other processes in Sample 3, as the pace and consequences inherent in MCI environments tended to overshadow interpersonal relations—at least in the short term.

Table 7
Team Process Correlations Over Time Per Sample

	Sample 1			Sample 2			Sample 3		
Previous (T-1) Processes	Trans	Action	Inter	Trans	Action	Inter	Trans	Action	Inter
Transition	.65	.63	.69	.33	.49	.25	.49	.44	.32
Action	.55	.56	.63	.36	.54	.29	.42	.46	.28
Interpersonal	.52	.53	.69	.40	.55	.45	.33	.32	.31

Note. Table values are correlations across episodes. Trans = Transition, Inter = Interpersonal.
Sample 1 ($N = 87$) and Sample 3 ($N = 449$) correlations all $p < .001$.
Sample 2 ($N = 62$) correlations: $>|.25|$, $p < .05$; $>|.33|$, $p < .01$; $>|.40|$, $p < .001$.

Team Process–Team Performance Over Time. Sample 3 also enabled us to test CATA-scored team process–criterion-related validity relationships. Notably, however, previous field research has typically correlated one-time measures of team processes with concurrent or lagged outcome measures (e.g., Mathieu & Schulze, 2006). Several student-sample investigations have correlated a small number of process and outcomes measures over episodes that each ranged from around 15 to 20 minutes in laboratory settings (e.g., Marks et al., 2005), to perhaps a few weeks in class-related projects (e.g., Mathieu & Schulze, 2006). However, as noted by Mathieu et al. (2020), such designs aggregate either team processes or performance that occurred over time (or both). Moreover, these designs either preclude the opportunity to model dynamic relations or limit them to the investigation of very few episodes. Sample 3 provides us with an opportunity to model dynamic relations as called for by Kozlowski (2015), Luciano, Mathieu, et al. (2018), and others.

Recall that earlier we established that team processes in this context could be meaningfully indexed in one-minute intervals. Accordingly, to preserve the temporal sequence of team processes-outcome relationships over time, we paired each 15-s outcome with team processes scored from the preceding minute. For instance, the team performance score for period 2–2:15 was paired with team process scores from 1 to 2 minutes. The 2:15–2:30 team performance score was paired with team processes from 1:15 to 2:15, and so on. The resulting descriptive statistics and correlations from these data are shown in Table 8. Notably, transition, action, and interpersonal processes all evidenced significant correlations with subsequent team performance ($r = .29$, .36, and .25, $p < .001$, respectively). Although these correlations are markedly lower than those reported in previous meta-analyses (LePine et al., 2008), they are not inflated by method effects. The growth modeling analyses described below accounts for the lack of independence of these scores over time (Bliese & Lang, 2016).

The design for the criterion-related analyses is a three-level mixed model. Level 1 represents a growth model with 1,078 team process–outcome pairings over time. Moreover, this is a dual-discontinuous change model (Bliese & Lang, 2016) whereby we control for both overall scenario time, and time since the first patient arrived in each team's area of operation. The former accounts for any general developmental process over the course of the scenario, whereas the latter is the more salient temporal factor for team operations. These temporal pairings are nested in three teams (Level 2) which were in turn nested in six scenarios (Level 3). Our analyses first model the dual-temporal factors introducing linear and quadratic trends for each. We then introduce team size and scenario type (coded: 1 = motor vehicle accident; 2 = mass shooting) as covariates, followed by the linear relationships with team processes. We then introduce team process by team time interactions, first with the linear parameter, and then with the quadratic parameter. Following the study by Marks et al. (2001), our expectations were that transition processes would be most predictive early and late during the performance period, and action processes would be most predictive during the middle period. We did not have any a priori expectations regarding interpersonal processes over time given that Marks et al. (2001) submitted that they could be salient both within and between episodes. We applied random coefficients growth modeling techniques in the form of the lme4 module in R (Bates et al., 2015) and employed a modeling-building approach proposed by Bliese and Ployhart (2002). These results are shown in Table 9. Because random coefficient models do not yield traditional R^2 estimates, we calculated a pseudo $\sim R^2$ using a method developed by Kreft and de Leeuw (1998).

Baseline analyses revealed that 37% of team performance variance occurred over time, 63% resided across teams, and 0% existed across scenarios. Regressing performance on to the temporal trends evidenced significant negative linear [$b = -.22$, $SE = .11$, $p < .05$] and positive quadratic effect [$b = .14$, $SE = .03$, $p < .01$] for scenario time, and a nonsignificant linear [$b = -.55$, $SE = .41$, ns] and positive quadratic effect [$b = .99$, $SE = .17$, $p < .01$] for team time (Model 1). Adding the two covariates in Model 2 revealed that performance was lower in the mass shooting scenarios [$b = -.14$, $SE = .05$, $p < .05$] and better when greater numbers of team members were present [$b = .05$, $SE = .01$, $p < .01$]. Introducing the team processes illustrated that only action processes [$b = .02$, $SE = .01$, $p < .01$] had a significant linear effect (Model 3). Adding

Table 8
Sample 3 Criterion-Related Validity Correlations

Variables	1	2	3	4	5	6	7	8	9	10	11
1. Scenario time[a]	—										
2. Team time[a]	.80	—									
3. Scenario type[b]	.08	−.07	—								
4. Team size[c]	.00	.35	−.09	—							
5. Team performance	−.08	.04	−.40	.46	—						
6. Transition processes	−.05	.19	−.12	.53	.29	—					
7. Action processes	−.01	.23	−.32	.56	.36	.60	—				
8. Interpersonal processes	−.05	.16	−.17	.48	.25	.41	.41	—			
9. % Women	−.11	−.23	−.14	−.07	.07	−.04	−.01	−.03	—		
10. Average age	.03	.08	.01	−.13	−.02	.05	.04	−.04	−.21	—	
11. Ethnicity Blau	−.08	.04	−.11	.55	.28	.27	.30	.28	.12	−.29	—
Team-Level											
Mean values	53.30	32.28	1.47	4.96	.71	1.60	1.88	1.06	.46	27.02	.42
Standard deviations	26.33	20.32	.50	3.43	.34	2.14	2.33	1.42	50.21	3.93	.20

Note. N =1,078 Observation Periods. Correlations > |.07|, $p < .05$; > |.08|, $p < .01$.
[a] Scaled as 15-second intervals.
[b] 1 = Motor vehicle accident, 2 = Mass shooting.
[c] Per observational period.

Table 9
Sample 3 Team Processes–Team Performance Growth Modeling Results

Predictors	DV = Team Performance				
	Model 1	Model 2	Model 3	Model 4	Model 5
1. Scenario linear time[a]	−.22 (.11)*	−.12 (.09)	−.12 (.09)	−.12 (.09)	−.13 (.09)
2. Scenario quadratic time[a]	.14 (.03)**	.13 (.03)**	.13 (.03)**	.13 (.04)**	.14 (.04)**
3. Team linear time[b]	−.55 (.41)	−.92 (.34)**	−.94 (.34)**	−.97 (.35)**	−.99 (.34)**
4. Team quadratic time[b]	.99 (.17)**	.97 (.16)**	.99 (.16)**	1.02 (.20)**	1.09 (.19)**
5. Scenario type[c]		−.14 (.05)*	−.14 (.05)*	−.14 (.05)*	−.14 (.05)*
6. Team size[d]		.05 (.01)**	.04 (.01)**	.04 (.01)**	.04 (.01)**
7. Transition processes			−.01 (.01)	−.03 (.02)	−.06 (.02)**
8. Action processes			.02 (.01)**	.05 (.02)**	.02 (.02)
9. Interpersonal Processes			.01 (.01)	−.01 (.01)	−.01 (.01)
10. Transition × Team Linear				.06 (.04)	.29 (.12)*
11. Action × Team Linear				−.09 (.04)*	.17 (.12)
12. Interpersonal × Team Linear				.03 (.03)	.03 (.03)
13. Transition × Team Quadratic					−.25 (.15)†
14. Action × Team Quadratic					−.27 (.15)†
15. Interpersonal × Team Quadratic					−.15 (.14)
$\Delta \sim R^2$	4.97**	25.29**	.37*	.33	1.87**
$\sim R^2$	4.97**	30.26**	30.63**	30.96**	32.83**

Note. Table values are unstandardized estimates, with standard errors within parentheses.
$N = 1{,}078$ observation periods nested in 3 teams and 6 scenarios.
[a] Scaled as standardized 15-s intervals.
[b] Scaled as 15-second intervals/100.
[c] 1 = Motor vehicle accident, 2 = Mass shooting.
[d] Per observational period.
† $p < .10$. * $p < .05$. ** $p < .01$.

the team processes by time interactions to the equation revealed that, again, only the action process by time interaction was significant [Model 4: $b = -.09$, $SE = .04$, $p < .05$]. A plot of this interaction (not shown) revealed that action processes were more strongly related to performance during initial team interactions and were less critical over time. Finally, as shown in Model 5, we then tested whether the three team processes by team quadratic time interactions would add significantly when introduced as a set. Although the set accounted for significant additional variance [$\Delta\chi^2_{(3)} = 21.89$, $p < .001$], the individual terms were not significant [interpersonal: $b = -.15$, $SE = .14$, ns; transition: $b = -.25$, $SE = .15$, $p = .09$; and action: $b = -.27$, $SE = .15$, $p = .07$]. Notably, however, eliminating the interpersonal process by quadratic time interaction from Model 5 rendered the latter two as significant [transition: $b = -.29$, $SE = .14$, $p < .05$ and action: $b = -30$, $SE = .15$, $p < .05$].

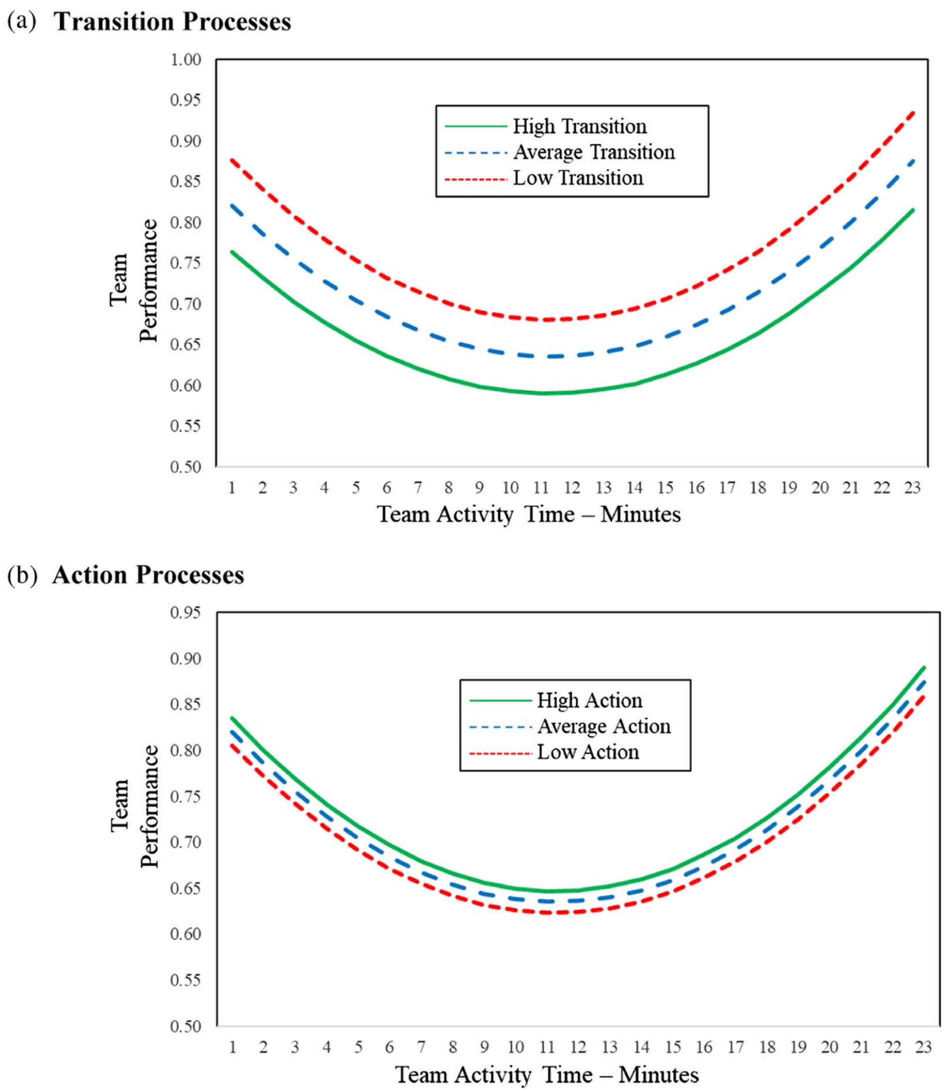
For illustrative purposes, Figure 4a and b shows the transition and action team processes' temporal interactions. In both cases, team processes were highly related to team performance as patients first arrived in their area of operation and waned over time. A noticeable upswing in the process–performance relationships, however, occurred around the average time that patients exited the team operational area (~15 minutes) and steadily increased throughout the longer performance periods. For both Figure 4a and b, the dashed (blue) curves represent the average process–performance relationship, the dotted (red) curves represent relatively low (−1 *SD*) process–performance relationships, and the solid (green) curves represent relatively high (+1 *SD*) process–performance relationship over time. Unexpectedly, stronger relationships were evident early and late in the performance periods by teams exhibiting relatively low transition processes and relatively high action processes. This belies predictions from the Marks et al.'s (2001) framework and the correlations reported above. However, upon further consideration and discussions with MCI trainers, there appears to be a good explanation for these findings.

Recall that these paramedic teams are operating in the context of an MTS. In dynamic and hectic MTS environments, it is important for higher-level MTS leadership to articulate an overall strategy and to facilitate coordination between teams (DeChurch & Marks, 2006). In this case, between-team coordination and directives from the incident commander should have accomplished important transition processes thereby liberating component teams to focus on action processes (i.e., treating patients). To the extent that the component teams were required to perform localized transition processes, it may have signaled less than effective MTS transition processes. Of course, these interpretations are post hoc and warrant further investigation. Regardless of the substantive relationships that these findings implicate, the fact that we were able to use the CATA scoring of team processes, together with a dynamic criterion measure, enabled us to test relationships in a fashion that has heretofore not been feasible.

Network Representations of Team Process. The final extension of our scoring protocol with Sample 3 was to illustrate dynamic network depictions of team processes, using communications emanating from each member of the MTS per minute. Crawford and LePine (2013) advanced a social network conception of team processes as a "set of ties or connections between members who interact to set goals, make plans, coordinate, help, and motivate each other. In other words, team members who indicate they set goals and make plans together, who monitor each other's progress and provide each other backup, or who work to manage each other's motivation and stress levels have a [multiple] teamwork tie[s]" (p. 34). To do so, we first evaluated the convergent validity of

Figure 4

Mass-Casualty Incident (MCI) Team Processes–Team Performance Relationships Over Time. (a) Transition Processes and (b) Action Processes



Note. See the online article for the color version of this figure.

hand-coded and CATA-scored team processes at the per minute *individual member level of analysis*. Then, we incorporated Bluetooth data from wearable sensors to index the senders and receivers of communications. Finally, multidimensional process networks were depicted per minute to showcase how CATA-scored team processes can be visualized.

Convergent and Discriminant Validity

The *individual-level* omega reliabilities for the hand-coded scores across 1,289 1-min episodes were as follows: (a) transition (3 indicators) $\omega = .22$, action (4 indicators) $\omega = .39$, and interpersonal (3 indicators) $\omega = .53$, whereas the corresponding omegas for the CATA scores were as follows: (a) transition (4 indicators) $\omega = .45$, action (5 indicators) $\omega = .91$, and interpersonal (4 indicators) $\omega = .69$. Here, again, the reliabilities are substantially lower than those found in Samples 1 and 2 but were derived from scores that are even more restricted than the team values noted above.

Table 10 shows descriptive statistics and correlations for Sample 3 member-level correlations. These findings show that the SME hand-coded scores correlate significantly with the CATA-scored transcripts for all three process dimensions: (a) transition: $r = .42$, $p < .001$; (b) action: $r = .38$, $p < .001$; and (c) interpersonal: $r = .29$, $p < .001$], averaging $r = .36$, $p < .001$.

We next reestimated the three-factor team processes model, which exhibited excellent fit indices [$\chi^2_{(6)} = 27.37, p < .01$; CFI = .980; SRMR = .020]. The factor loadings for this model are also shown in Figure 5. As shown, all indicators loaded significantly on their intended latent variables, ranging from $\Gamma = .38, p < .001$ to $\Gamma = .85$, $p < .001$, averaging $\Gamma = .62$, $p < .001$. The correlations among the latent variables were all significant as follows: (a) transition–action: $\Phi = .68$, $p < .001$; (b) transition–interpersonal:

Table 10
Sample 3 Individual Level Descriptive Statistics and Correlations

Dimensions	1	2	3	4	5	6
1. Transition_H	(.22)					
2. Action_H	.17	(.39)		H-C: $r = .36$, $p < .001$		
3. Interpersonal_H	.08	.13	(.53)			
4. Transition_C	.42	.27	.14	(.45)		
5. Action_C	.24	.38	.16	.44	(.91)	
6. Interpersonal_C	.22	.29	.29	.30	.26	(.69)
Individual-Level						
Means	.76	3.57	.21	1.58	2.00	1.10
Standard Deviations	1.16	2.91	.66	1.95	2.17	1.44

Note. N = 1,289 Individual Minute Episodes. Correlations: > |.05|, $p < .05$; > |.07|, $p < .01$; > |.09|, $p < .001$.

H = SME Hand-Coded, C = CATA-Scored.

Rectangles highlight scale convergent validities with their average provided in the summary on the right.

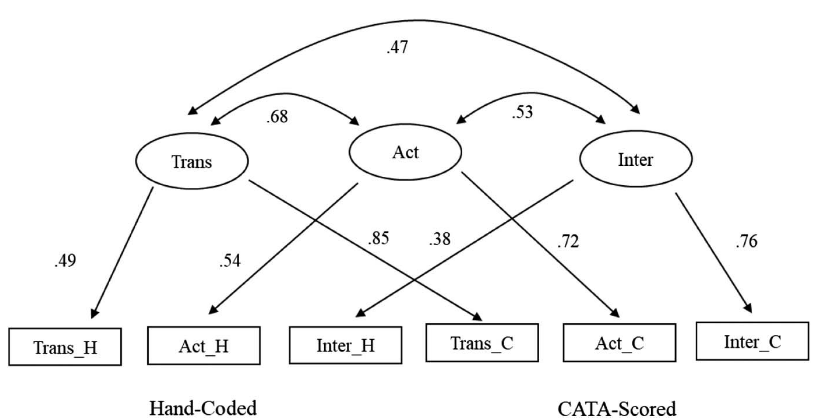
Numbers in parentheses represent reliability diagonal with omega for Hand-Coded and CATA-scored.

$\Phi = .47$, $p < .001$; and (c) action–interpersonal: $\Phi = .53$, $p < .001$], averaging $\Phi = .56$, $p < .001$. We fit a two-factor model constraining the correlation between the latent transition and action processes to 1.0, which revealed an acceptable fit [$\chi^2_{(7)} = 85.22$, $p < .001$; CFI = .926; SRMR = .035], but was significantly worse [$\Delta\chi^2_{(1)} = 57.86$, $p < .001$] than the three-factor model. Again, these findings provide evidence of the discriminant validity among the process dimensions indexed at the member level of analysis.

Use of Bluetooth Data

The ability to index team processes emanating from each member enables one to construct time-sensitive networks of team interactions. This would be relatively easy to accomplish if members, for example, communicated through email or text or were clearly colocated such as in a conference room. But, it is more difficult to discern to whom a given individual is directing communications in a dynamic environment, such as an MCI, where the paramedics

Figure 5
Sample 3 Individual-Level Three-Factor Model Results



Note. Trans = Transition processes; Act =Action processes; Inter = Interpersonal processes.
H = Hand-coded transcripts; C = Computer-aided text analysis.
Standardized Factor Loadings shown, $p < .001$.
Values from N = 1,289 individual minute episodes.

move around frequently. Moreover, although deployed as three different teams, our sample paramedics functioned as a small MTS (Luciano, DeChurch, & Mathieu, 2018) and often coordinated their activities with members of other teams as they operated in space and physically moved patients from one stage to another. At issue, then, is determining on a minute by minute basis, to whom each paramedic was communicating.

Luciano, Mathieu, et al. (2018), and Müller et al. (2019), both suggested that multimodal streams of information might be combined to index members' interactions. For instance, Bluetooth data from wearable sensors can permit one to determine which members are within speaking range of others (cf., Kim et al., 2012; Matusik et al., 2019). Based on a qualitative analysis of our MCI environment, we determined that members within approximately 10 feet were able to hear one another.[5] Therefore, we first established minute by minute proximity networks based on paramedics' Bluetooth signatures, and then overlaid their respective process communications. Notably, our approach uses members' proximity data simply to inform whether individuals were likely to be privy to others' communications, not as a proxy for substantive relationships. As such, we CATA scored each member's communications in terms of Marks et al.'s three higher-order dimensions as outward ties linked to other members within hearing range during that minute.

Multidimensional Team Process Networks. The steps above yield multidimensional process networks, per minute, such as the ones shown in Figure 6. The networks shown in Figure 6 are illustrative examples of the multidimensional relationships at different phases of an MCI. The lines represent continuously scored transition, action, and interpersonal process ties in each of three different 1-minute periods. Thickness of the lines represents greater intensity of processes. For instance, the Phase 1 set of networks (from the fourth minute of a mass shooting scenario) illustrate relatively sparse processes were mainly occurring within teams with limited cross-team processes. The incident commander was directing his/her transition and action processes strictly to the triage team, and the patterns of the three process networks (i.e., transition, action, and interpersonal) were relatively similar. The surprising value of action processes during this phase, suggested by the criterion-related analyses reported above, is also evident in the density of the lines in the corresponding network. Overall, these network images are representative of operations during the initial triage phase of an MCI.

The Phase 2 set of networks shown in Figure 6 are from minute 15 of the same scenario and exhibit much denser networks and a higher percentage of cross-team processes. Transition processes were more pronounced and focused on the treatment team, whereas action processes were concentrated on the transport team. Interpersonal processes were greater than in Phase 1 and balanced throughout the network, but less present than transition or action processes. These patterns are representative of times when the MCI was focusing on stabilizing and reassessing patients while sequencing and transferring them from treatment to transport. The Phase 3 set of networks are from minute 26 of the scenario when the action processes were concentrated mostly on the transport team. Interestingly, transition processes were mostly between triage and treatment members at this stage. Anecdotally, our experience has been that because the work of these members was ending by this time, they would begin to engage in spontaneous after-action review style conversations—the reflection portion of Marks et al.'s transition processes. Interpersonal processes often accompany these discussions as shown in the last network.

Discussion

Despite decades of calling for incorporation of temporal perspectives into organizational research, there has been limited progress along these lines. A key impediment to these important advancements has been the development of construct valid unobtrusive measures of collective constructs which are necessary to align theory and measurement. Thus, we developed a CATA protocol for indexing dynamic collective constructs based on participants' communications. We featured Marks et al.'s (2001) taxonomy of team processes and employed an iterative content validity-based approach to develop dictionaries of their 10 lower-order and 3 higher-order dimensions. We then further refined those dictionaries and established their convergent, discriminant, and predictive validity using three different samples. Below, we further discuss the contributions of our study and how they can be leveraged to enable the study of any dynamic collective constructs over time.

Implications

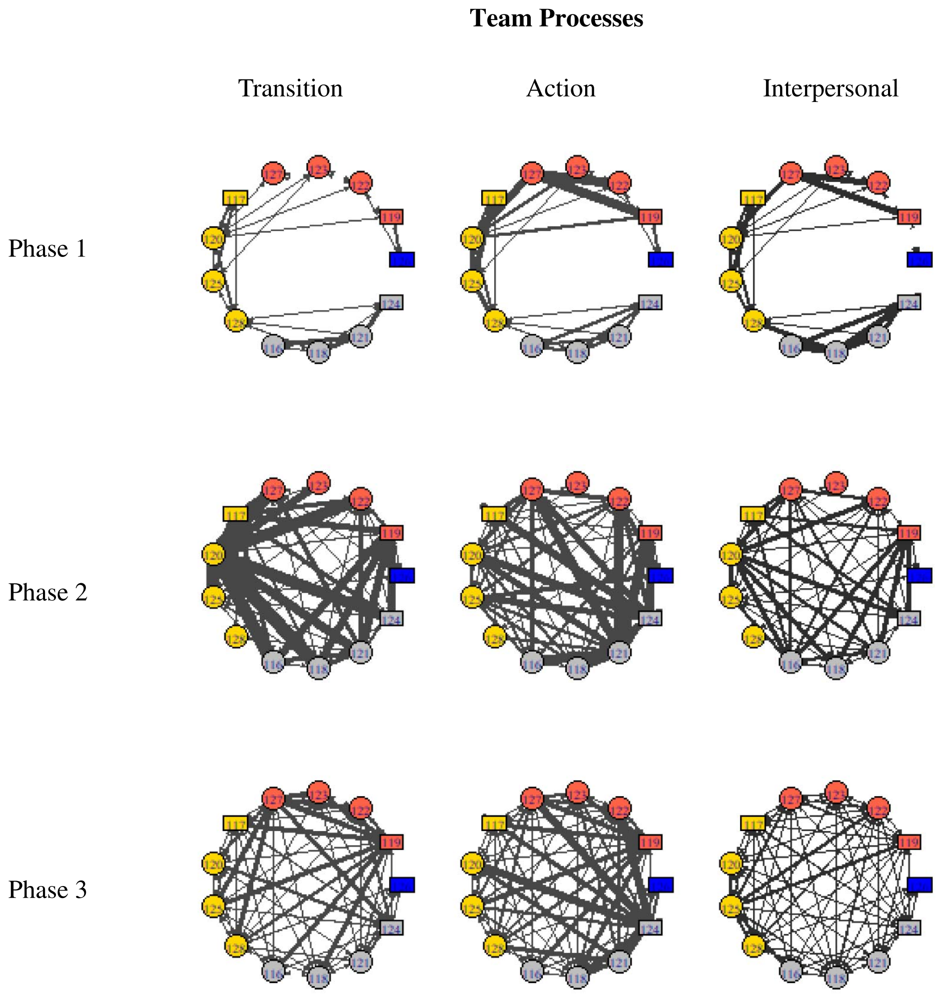
This study fulfills its overarching purpose of illustrating a method to use analyses of individuals' communications as construct valid unobtrusive measures of collective constructs occurring dynamically over time. Thus, we contribute both an approach for reliable and valid CATA dictionary development based on participants' communications and a specific CATA protocol for a popular team process framework. Our approach is uniquely well suited to assess the Marks et al.'s (2001) framework as it captures the multiple dimensions of team processes whenever they occur, across multiple tasks and episodes.

Using transcripts of teammates' communications from two laboratory samples, we demonstrated the convergent validity of CATA-scored transcripts with SMEs' hand-coding of those transcripts, and with other SMEs' BARS ratings of members' interactions based on video recordings. In addition, using a sample of paramedics performing a series of MCI scenarios, we found further convergent validity between the CATA and SME hand-coded scorings of transcripts. Thus, we contribute rigorous support for convergent validity for our team process dictionaries against the high fidelity "gold standard" of SME hand-coding across three samples using controlled laboratory data from smaller teams, and dynamic real-world tasks from MTSs.

Moreover, we unearthed complex temporally dependent process–performance relationships that would not have been evident using traditional measures and designs. We also illustrated how team processes emanating from individuals could be scored using CATA and used to depict multidimensional networks of relations. This microdynamic approach to indexing team processes that we illustrated with Sample 3 can finally enable the testing of dynamic theories of team processes that have long been called for (cf., Cronin

[5] In a pilot study, we found significant convergent validity ($r = .47$, $p < .001$) between this method of indexing member interactions and observers' first-hand coding of team interactions.

Figure 6
Sample 3 Multidimensional Team Process Networks Over Time



Note. Lines depict directed team processes with greater thickness representing greater intensity. Red symbols = Triage; yellow symbols = Treatment; gray symbols = Transport; circles = Members; rectangles = Leaders; blue rectangle = Incident Commander. See the online article for the color version of this figure.

et al., 2011; Marks et al., 2001; McGrath, 1991). Liberated from constraints and threats to validity such as survey fatigue, testing, or instrumentation effects, CATA-scored team processes can be used in growth-modeling (e.g., Bliese & Lang, 2016), experience sampling (Beal, 2015), and perhaps event-based (Morgeson et al., 2015) designs to model far more nuanced and complex relations than have heretofore been possible (Mathieu & Luciano, 2019). Notably, the longitudinal criterion-related validity evidence that we illustrated from Sample 3 offers an early indication of what lies ahead, in that, the relationships it revealed differed markedly from those observed in previous work using more traditional designs and analyses (cf., LePine et al., 2008).

CATA scoring of collective constructs also suggests important practical implications for researchers in terms of time and resources. For instance, for the Sample 1 data, it required several weeks of training and watching team recordings for our SMEs to generate the BARS ratings. The other two SMEs who completed the hand-coding of transcripts took weeks to develop a common scoring protocol, and then approximately 3 months to complete their hand-coding. Conversely, once the generic and supplemental dictionaries were developed, the CATA scoring of those same team interactions took less than 5 seconds to complete. This is not to suggest that the CATA scoring is a panacea or necessarily better than other methods, but it does enable the modeling of dynamic team process relations over time.

Furthermore, the rapid processing and scalability (post dictionary development) unlocks the potential for data-informed decision making *during* team performance episodes and for immediate postperformance team debriefs. Team debriefs "are designed to improve teamwork processes by engaging team members in active learning throughout the learning cycle" (Lacerenza et al., 2018, p. 525). During team debriefs, members reflect on a performance episode or experience; however, team members may not always understand what caused the team process or performance to succeed or fail within this episode. Tannenbaum and Cerasoli (2013) note that "multimedia aids may be one way of building in structure and guidance" (p. 240). For instance, Stephanian et al. (2015) find that technology, such as video, can help to provide structured inputs for team debriefs. Similarly, incorporating this CATA-generated data can help to provide this structure when identifying which parts of team process occurred (or broke down) during a given performance episode. Moreover, this data can help point teams and leaders to specific examples of problems and areas to improve as well as confirm successes. In addition to team debriefs, using CATA-generated data could be similarly useful in team training by highlighting specific behaviors and processes for both leaders and teams.

Limitations and Future Directions

Our article unlocks the potential to explore numerous research questions related to multilevel and dynamic processes by testing temporal-based theories of team processes, in such a way that better aligns measurement and theory. Nonetheless, it is not without its limitations. Below, we discuss the limitations of the current work and offer specific areas for future consideration.

There are several practical limitations regarding the use of CATA. We previously mentioned that once completed CATA using LIWC took less than 5 seconds to complete as compared to the month's long effort of training coders and actual coding. However, this glosses over the fact that it takes considerable time and effort to create the first iteration and to validate CATA dictionaries—as well as the importance of creating context-specific supplements. However, these are largely "upfront" costs with limited maintenance costs and allow for substantial time savings on the back end. Furthermore, building and validating CATA dictionaries for one-off endeavors may not necessarily justify the costs, whereas the costs would easily be justified in instances of multiple applications.

Another limitation with the use of CATA is that it only captures the variance available in text, which is often derived from verbal communication. This is especially notable for interpersonal process, where the same words and phrases can bear completely different meanings depending on the context and the tone of the voice (e.g., sarcasm). In addition, there is a need to either acquire text data or to transcribe audio which can be costly. Furthermore, many well-orchestrated team processes are executed with simple hand gestures, short phrases, and implicit forms of coordination. Well-functioning teams performing complex tasks—such as surgery and cockpit crews—often have an eerie quiet to them. Nevertheless, we suggest that team members' communications are more revealing than not in our study contexts, and that our dictionaries provide a foundation for further development and application.

In addition, across our three samples, lower-order dimensions did not offer the same fidelity as the higher-order dimensions. In many ways, this can be somewhat disheartening when considering the time-intensive efforts to build these dictionaries. However, by starting with the lower-order constructs that make up the higher-order constructs we were able to better triangulate on the higher-order constructs. Furthermore, from a bandwidth standpoint (Cronbach & Gleser, 1957), given our relatively broad criterion of team performance (i.e., dynamic patient sequencing) aligning the breadth versus specificity of the predictors and outcomes in the future applications may highlight differential utility of lower- and higher-order team processes.

Another limitation in our current application is that we utilized a relatively primitive application of CATA in that we simply feature word counts, expanded only to word counts with networks in Sample 3. One fruitful area for the future research is to combine CATA with other modalities to generate dynamic multidimensional network depictions of team processes (Crawford & LePine, 2013) or other organizational constructs such as collaboration or conflict (Park, Mathieu, et al., 2020). For instance, our Sample 3 would yield substantial data with over 450 networks (i.e., three networks per 153 minutes = 459) that could be analyzed using a variety of powerful techniques. For instance, Leenders et al. (2016) advanced a longitudinal social network methodology that "is minimally characterized by the time at which the interaction was initiated, the team member who initiated it, and the team member(s) who were the recipients" (pp. 97–98). They describe how such analyses can be used to identify *sequential structure signatures*, which are hypothesized patterns of interactions over time. Elsewhere, exponential random graph modeling (ERGM) enables scholars to derive "graph motifs," which are representations of the nature of members' interactions, and Sewell and Chen (2016) have demonstrated how ERGM models can be used to model dynamic properties over time. Other advancements have included multilevel ERGM (MERGM), which can simultaneously analyze networks that are nested in the traditional sense, such as teams in MTS (cf., Slaughter & Koehly, 2016; Zappa & Robins, 2016). The summary point, however, is that CATA-scored team processes, combined with a means by which intermember communications can be specified (e.g., using Bluetooth signatures), enable the construction and modeling of dynamic multidimensional networks over time, suitable for a wide variety of research designs and analyses (Park, Grosser et al., 2020).

Our current application is also limited by our focus on one framework using positive terms. Future work should further expand and refine our dictionaries, with additional supplements for other task contexts or potentially by exploring negative terms. As noted earlier, our work only features the positive terms for the Marks et al.'s (2001) dimensions. We had initially sought to generate negative terms, such as "argue" for conflict management or "threw-together" for strategy formulation/planning. However, our initial investigations did not find that they correlated negatively with words intended to be positive indicators. This may be attributable to the fact that the Marks et al.'s (2001) dimensions are labeled in inherently positive terms, or perhaps that positive and negative instances of team processes may not be on one continuum (e.g., motivate vs. depress) in the same manner as positive and negative affectivity (Watson et al., 1988). Moreover, there is an open question as to whether talking about negative team processes is, in fact, a positive team process. In other words, discussing members' interpersonal conflicts may be explosive—or the path toward resolution—yet, word counts would have difficulty differentiating the two exchanges. Therefore, future researchers should investigate what seemingly negative words signal in terms of team processes (cf., Short et al., 2018).

In addition, discerning to whom a member's communications are directed represents a potential challenge. This may be relatively easy to distinguish in directed chats or emails, or when members are collocated in a small space. But if members are mobile, it might require additional data such as the position information we leveraged via Bluetooth signals in Sample 3 (Luciano, Mathieu, et al., 2018). Yet, distinctions may still not be clear-cut. For instance, should individuals cc'ed on an email be considered full- or partial-recipients? How should bystanders who overhear communications not intended for them be treated? How should communications directed toward focal individuals who are not paying attention or who are otherwise distracted be scored? Such questions are not unique to CATA, but they do force investigators to consider these matters as they operationalize team process patterns.

Finally, our studies generated the raw data for our analysis from transcriptions of audio-recorded team communications. However, other sources of data (e.g., chat/texts, emails, and social media postings) also generate the text necessary for CATA processing. This raises the question as to whether the "medium matters" in teams' communications. The virtual team's literature suggests that the use of different mediums (e.g., chat, threaded discussion, and videoconferencing) may enhance or impoverish different types of team processes (Kirkman & Mathieu, 2005). Thus, technology offers promise not only in terms of how best to capture and process communications data but also may have implications in terms of the quality and value of such communications. This opens a variety of substantive and methodological questions for the future research.

Additional Considerations

We see great promise in CATA and other streaming data measures of organizational constructs. But, these newer technologies resurrect old, and present some new, challenges to construct validity. Below, we highlight important considerations related to accuracy, feasibility, completeness, and ethics. Each of these factors warrants detailed consideration with regards to the specific construct elements, additional measurement features, research question, and study context (see Luciano, Mathieu, et al., 2018).

Although our illustrations featured team processes, similar applications could be developed for other individual and collective constructs in applied psychology (e.g., culture, collaboration, and shared leadership) as well as more macro collective phenomenon (e.g., human capital emergence—Ployhart & Moliterno, 2011). When doing so, researchers need to consider the appropriate strategy that suits the topic area. As shown in our illustration, deductive style CATA approaches are suitable for instances where there is an existing theoretical framework and well-articulated constructs. In contrast, inductive CATA approaches are better suited for nascent areas where the construct domain has yet to be clearly articulated (Short et al., 2018).

We note that CATA comes in various approaches ranging from simple word counts, pronoun comparisons (e.g., singular vs. plural rates), and sentiment analysis, to complex natural language processing. The former varieties lose much of the nuance embedded in members' communications, whereas more sophisticated techniques are complex and time consuming to implement. We sought to establish a "middle ground" where content valid generic dictionaries could be supplemented with context-specific additions and adjustments to provide a scoring protocol that is relatively easy to implement yet yields informative scores. Whereas we grounded our work in a well-established taxonomy of team processes (Marks et al., 2001) and provided evidence of content and construct validity, we readily admit that traditional qualitative analyses of communications provide a richer understanding of the nuances of team interactions (e.g., Meinecke et al., 2017). For instance, although SMEs evidenced good consistency classifying words into the three higher-order Marks et al.'s dimensions, they had difficulty differentiating words into the 10 lower-order dimensions. Many words were classified differently or seen as overlapping those more specific dimensions. Similar correlations have been observed with survey measures, and we echo the sentiments expressed by others that the purposes of the investigation should guide whether one seeks to index lower- or higher-order constructs (e.g., LePine et al., 2008; Mathieu et al., 2020). Yet, it is fair to say that the CATA dictionaries approach may be limited if nuanced distinctions are desired. We fully anticipate that as natural language processing algorithms evolve, so too, will the sophistication of CATA techniques and perhaps their ability to differentiate the lower-order dimensions.

Our use of CATA also brings to the forefront temporal questions which have been noted by Marks et al. (2001), Kozlowski (2015), and Luciano, Mathieu, et al. (2018). For instance, Kozlowski (2015) advocates indexing team processes at the highest rate possible to yield detailed change patterns. However, our analysis suggested that small samples of communications are less reliable—particularly when communications were limited. Luciano, Mathieu, et al. (2018) suggested that although larger samples of spoken or written words will likely yield more reliable and stable measures, they may obscure important nuances or dynamics inherent in the different data streams. Finding the optimal temporal "chunking" or unitization rates is likely be an iterative process considering the nature of the construct in question, study context/logistics, and one's theory of time and events. Thus, CATA indexing of team processes is best determined using a mixed-method measurement fitting process whereby the temporal rhythms of team activities are identified and used to guide the frequency of measurement and aggregation.

The use of individual communication streams raises some issues concerning ethical and privacy implications. Our participants all willingly provided informed consent to video and/or audio tape their interactions, which were approved by the university Institutional Review Boards. However, not all employees in work organizations are afforded the same options, as several must agree to be monitored as a condition of employment, and others are monitored unwittingly. There can certainly be different perspectives on these practices. For instance, being able to monitor the physical condition, oxygen intake rate, communications, and surrounding physical conditions of firefighters in burning structures enable incident commanders to optimally deploy, position, and withdraw team members to enhance their safety and effectiveness (Devine et al., 2018). Combining real-time aircraft status and position data with analyses of cockpit–air traffic control communications, particularly under emergency circumstances, may be beneficial to all involved (Rusko & Finke, 2016). But whether, for instance, recording accounting team members performing normal audits, sales team interactions, or idle conversation among virtual team members before the formal start of a meeting, crosses the line of privacy is another question (Moussa, 2015). In addition, how such data should be maintained and available for reproducibility or transparency purposes is also subject to debate. And finally, whether such data are subject to subpoena is an open ethical and legal question (Wolf et al., 2015). Of course, none of the cautions mentioned here are unique to CATA. Data collections done using surveys, observations, interviews, and other digital traces are subject to the same issues. But with the ubiquity of recording devices in modern-day organizations, we encourage greater open dialogue.

Conclusion

In summary, we believe that CATA of team and other organizational constructs represent a viable, construct valid means by which dynamic processes can be indexed and modeled. Our current initiative provides a means to better align theory and measurement, balancing the benefits of extracting a rich understanding of members' interactions revealed through their words, with the desire to employ a reliable, valid, and scalable method to index such communications. Thus, we provide an infrastructure to enable the testing of complex theories and generate enhanced understanding of team processes and other dynamic constructs. Our approach opens up important avenues for the future research that truly captures the "dynamics" of team dynamics—yet also raises intriguing theoretical and methodological considerations.

References

Ancona, D. G., Okhuysen, G. A., & Perlow, L. A. (2001). Taking time to integrate temporal research. *Academy of Management Review*, *26*(4), 512–529. https://doi.org/10.5465/amr.2001.5393887

Anderson, J. C., & Gerbing, D. W. (1991). Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *Journal of Applied Psychology*, *76*(5), 732–740. https://doi.org/10.1037/0021-9010.76.5.732

Ashkanasy, N. M., Humphrey, R. H., & Huy, Q. N. (2017). Integrating emotions and affect in theories of management. *Academy of Management Review*, *42*(2), 175–189. https://doi.org/10.5465/amr.2016.0474

Bales, R. F. (1950). A set of categories for the analysis of small group interaction. *American Sociological Review*, *15*, 257–263. https://doi.org/10.2307/2086790

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*(4), 396–403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Beal, D. J. (2015). ESM 2.0: State of the art and future potential of experience sampling methods in organizational research. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*(1), 383–407. https://doi.org/10.1146/annurev-orgpsych-032414-111335

Bedwell, W. L. (2019). Adaptive team performance: The influence of membership fluidity on shared team cognition. *Frontiers in Psychology*, *10*, 2266. https://doi.org/10.3389/fpsyg.2019.02266

Bedwell, W. L., Wildman, J. L., DiazGranados, D., Salazar, M., Kramer, W. S., & Salas, E. (2012). Collaboration at work: An integrative multilevel conceptualization. *Human Resource Management Review*, *22*(2), 128–145. https://doi.org/10.1016/j.hrmr.2011.11.007

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Bliese, P. D., & Lang, J. W. (2016). Understanding relative and absolute change in discontinuous growth models: Coding alternatives and implications for hypothesis testing. *Organizational Research Methods*, *19*(4), 562–592. https://doi.org/10.1177/1094428116633502

Bliese, P. D., & Ployhart, R. E. (2002). Growth modeling using random coefficient models: Model building, testing, and illustrations. *Organizational Research Methods*, *5*(4), 362–387. https://doi.org/10.1177/109442802237116

Brauner, E., Boos, M., & Kolbe, M. (2018). *The Cambridge handbook of group interaction analysis*. Cambridge University Press. https://doi.org/10.1017/9781316286302

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. https://doi.org/10.1037/h0046016

Christianson, M. K. (2018). Mapping the terrain: The use of video-based research in top-tier organizational journals. *Organizational Research Methods*, *21*(2), 261–287. https://doi.org/10.1177/1094428116663636

City of Phoenix. (2000, March). *Phoenix regional standard operating procedures*. M. P. 203.01. https://www.phoenix.gov/firesite/Documents/074745.pdf

Colquitt, J. A., Sabey, T. B., Rodell, J. B., & Hill, E. T. (2019). Content validation guidelines: Evaluation criteria for definitional correspondence and definitional distinctiveness. *Journal of Applied Psychology*, *104*(10), 1243–1265. https://doi.org/10.1037/apl0000406

Cote, J. A., & Buckley, M. R. (1987). Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *JMR, Journal of Marketing Research*, *24*(3), 315–318. https://doi.org/10.1177/002224378702400308

Crawford, E. R., & LePine, J. A. (2013). A configural theory of team processes: Accounting for the structure of taskwork and teamwork. *Academy of Management Review*, *38*(1), 32–48. https://doi.org/10.5465/amr.2011.0206

Cronbach, L. J., & Gleser, G. C. (1957). *Psychological tests and personnel decisions*. University of Illinois Press.

Cronin, M. A., Weingart, L. R., & Todorova, G. (2011). Dynamics in groups: Are we there yet? *The Academy of Management Annals*, *5*(1), 571–612. https://doi.org/10.5465/19416520.2011.590297

DeChurch, L. A., & Marks, M. A. (2006). Leadership in multiteam systems. *Journal of Applied Psychology*, *91*(2), 311–329. https://doi.org/10.1037/0021-9010.91.2.311

Devine, M., Bond, R., Simms, V., Boyce, K., & Kerr, D. (2018). *Mapping the health, safety and wellbeing challenges of firefighting to wearable devices*. Paper presented at the British HCI Conference, Belfast, UK.

Dionne, S. D., Gupta, A., Sotak, K. L., Shirreffs, K. A., Serban, A., Hao, C., . . . Yammarino, F. J. (2014). A 25 year perspective on levels of analysis in leadership research. *The Leadership Quarterly*, *25*(1), 6–35. https://doi.org/10.1016/j.leaqua.2013.11.002

Fischer, U., McDonnell, L., & Orasanu, J. (2007). Linguistic correlates of team performance: Toward a tool for monitoring team functioning during space missions. *Aviation, Space, and Environmental Medicine*, *78*(5), B86–B95.

Hart, R. P., Carroll, C. E., & Spiars, S. (2015). *Diction 7: A text analysis program*. Digitext.

Hitt, M. A., Beamish, P. W., Jackson, S. E., & Mathieu, J. E. (2007). Building theoretical and empirical bridges across levels: Multilevel research in management. *Academy of Management Journal*, *50*(6), 1385–1399. https://doi.org/10.5465/amj.2007.28166219

Honnibal, M. (2016). *Syntaxnet in context: Understanding google's new tensorflow nlp model*.

Humphrey, S. E., & LeBreton, J. M. (2019). *The handbook of multilevel theory, measurement, and analysis*. American Psychological Association. https://doi.org/10.1037/0000115-000

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organizations: From I-P-O models to IMOI models. *Annual Review of Psychology*, *56*, 517–543. https://doi.org/10.1146/annurev.psych.56.091103.070250

Kim, T., McFee, E., Olguin, D. O., Waber, B., & Pentland, A. (2012). Sociometric badges: Using sensor technology to capture new forms of collaboration. *Journal of Organizational Behavior*, *33*(3), 412–427. https://doi.org/10.1002/job.1776

Kirkman, B. L., & Mathieu, J. E. (2005). The dimensions and antecedents of team virtuality. *Journal of Management*, *31*(5), 700–718. https://doi.org/10.1177/0149206305279113

Klein, K. J., & Kozlowski, S. W. (2000). *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*. Jossey-Bass.

Kozlowski, S. W. J. (2015). Advancing research on team process dynamics: Theoretical, methodological, and measurement considerations. *Organizational Psychology Review*, *5*(4), 270–299. https://doi.org/10.1177/2041386614533586

Kreft, I. G., & De Leeuw, J. (1998). *Introducing multilevel modeling*. SAGE Publications. https://doi.org/10.4135/9781849209366

Lacerenza, C. N., Marlow, S. L., Tannenbaum, S. I., & Salas, E. (2018). Team development interventions: Evidence-based approaches for improving teamwork. *American Psychologist*, *73*(4), 517–531. https://doi.org/10.1037/amp0000295

Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological Methods*, *7*(2), 228–244. https://doi.org/10.1037/1082-989X.7.2.228

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 139–174. https://doi.org/10.2307/2529310

Leenders, R. T. A., Contractor, N. S., & DeChurch, L. A. (2016). Once upon a time: Understanding team processes as relational event networks. *Organizational Psychology Review*, *6*(1), 92–115. https://doi.org/10.1177/2041386615578312

Lehmann-Willenbrock, N., & Allen, J. A. (2018). Modeling temporal interaction dynamics in organizational settings. *Journal of Business and Psychology*, *33*(3), 325–344. https://doi.org/10.1007/s10869-017-9506-9

LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: Tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology*, *61*(2), 273–307. https://doi.org/10.1111/j.1744-6570.2008.00114.x

Lockwood, C., Giorgi, S., & Glynn, M. A. (2019). "How to do things with words": Mechanisms bridging language and action in management research. *Journal of Management*, *45*(1), 7–34. https://doi.org/10.1177/0149206318777599

Luciano, M. M., DeChurch, L. A., & Mathieu, J. E. (2018). Multiteam systems: A structural framework and meso-theory of system functioning. *Journal of Management*, *44*(3), 1065–1096. https://doi.org/10.1177/0149206315601184

Luciano, M. M., Mathieu, J. E., Park, S., & Tannenbaum, S. I. (2018). A fitting approach to construct and measurement alignment: The role of big data in advancing dynamic theories. *Organizational Research Methods*, *21*(3), 592–632. https://doi.org/10.1177/1094428117728372

Marks, M. A., DeChurch, L. A., Mathieu, J. E., Panzer, F. J., & Alonso, A. (2005). Teamwork in multiteam systems. *Journal of Applied Psychology*, *90*(5), 964–971. https://doi.org/10.1037/0021-9010.90.5.964

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review*, *26*(3), 356–376. https://doi.org/10.5465/amr.2001.4845785

Mathieu, J. E., & Chen, G. (2011). The etiology of the multilevel paradigm in management research. *Journal of Management*, *37*(2), 610–641. https://doi.org/10.1177/0149206310364663

Mathieu, J. E., Gallagher, P. T., Domingo, M. A., & Klock, E. A. (2019). Embracing complexity: Reviewing the past decade of team effectiveness research. *Annual Review of Organizational Psychology and Organizational Behavior*, *6*, 17–46. https://doi.org/10.1146/annurev-orgpsych-012218-015106

Mathieu, J. E., Hollenbeck, J. R., Knippenberg, D. V., & Ilgen, D. R. (2017). A century of work teams in the journal of applied psychology. *Journal of Applied Psychology*, *102*(3), 452–467. https://doi.org/10.1037/apl0000128

Mathieu, J. E., & Luciano, M. M. (2019). Multilevel emergence in work collectives. In S. E. Humphrey & J. M. LeBreton (Eds.), *The handbook for multilevel theory, measurement, and analysis* (pp. 163–186). American Psychological Association. https://doi.org/10.1037/0000115-008

Mathieu, J. E., Luciano, M. M., D'Innocenzo, L., Klock, E. A., & LePine, J. A. (2020). The development and construct validity of a team processes survey measure. *Organizational Research Methods*, *23*, 399–431. https://doi.org/10.1177/1094428119840801

Mathieu, J. E., & Schulze, W. (2006). The influence of team knowledge and formal plans on episodic team process–performance relationships. *Academy of Management Journal*, *49*(3), 605–619. https://doi.org/10.5465/amj.2006.21794678

Mathieu, J. E., & Taylor, S. R. (2006). Clarifying conditions and decision points for mediational type inferences in organizational behavior. *Journal of Organizational Behavior*, *27*, 1031–1056. https://doi.org/10.1002/job.406

Mathieu, J. E., Wolfson, M. A., & Park, S. (2018). The evolution of work team research since Hawthorne. *American Psychologist*, *73*(4), 308–321. https://doi.org/10.1037/amp0000255

Matusik, J. G., Heidl, R., Hollenbeck, J. R., Yu, A., Lee, H. W., & Howe, M. (2019). Wearable bluetooth sensors for capturing relational variables and temporal variability in relationships: A construct validation study. *Journal of Applied Psychology*, *104*(3), 357–387. https://doi.org/10.1037/apl0000334

Maynard, M. T., Gilson, L. L., & Mathieu, J. E. (2012). Empowerment—fad or fab? A multilevel review of the past two decades of research. *Journal of Management*, *38*(4), 1231–1281. https://doi.org/10.1177/0149206312438773

McDonald, R. P. (1999). *Test theory: A unified approach*. Lawrence Erlbaum.

McElroy, J. C., & Downey, H. K. (1982). Observation in organizational research: Panacea to the performance-attribution effect? *Academy of Management Journal*, *25*(4), 822–835. https://doi.org/10.5465/256101

McGrath, J. E. (1991). Time, interaction, and performance (TIP): A theory of groups. *Small Group Research*, *22*, 147–174. https://doi.org/10.1177/1046496491222001

McKenny, A. F., Short, J. C., & Payne, G. T. (2012). Using computer-aided text analysis to elevate constructs: An illustration using psychological capital. *Organizational Research Methods*, *16*(1), 152–184. https://doi.org/10.1177/1094428112459910

Meinecke, A. L., Lehmann-Willenbrock, N., & Kauffeld, S. (2017). What happens during annual appraisal interviews? How leader–follower interactions unfold and impact interview outcomes. *Journal of Applied Psychology*, *102*(7), 1054–1074. https://doi.org/10.1037/apl0000219

Mistovich, J. J., Karren, K. J., & Hafen, B. (2013). *Prehospital emergency care* (10th ed.). Prentice-Hall.

Mitchell, T. R., & James, L. R. (2001). Building better theory: Time and the specification of when things happen. *Academy of Management Review*, *26*(4), 530–547. https://doi.org/10.5465/amr.2001.5393889

Morgeson, F. P., Mitchell, T. R., & Liu, D. (2015). Event system theory: An event-oriented approach to the organizational sciences. *Academy of Management Review*, *40*(4), 515–537. https://doi.org/10.5465/amr.2012.0099

Moussa, M. (2015). Monitoring employee behavior through the use of technology and issues of employee privacy in America. *SAGE Open*, *5*(2), 1–13. https://doi.org/10.1177/2158244015580168

Müller, J., Fàbregues, S., Guenther, E. A., & Romano, M. J. (2019). Using sensors in organizational research—clarifying rationales and validation challenges for mixed methods. *Frontiers in Psychology*, *10*(1188), 1–14. https://doi.org/10.3389/fpsyg.2019.01188

Park, S., Grosser, T. J., Roebuck, A. A., & Mathieu, J. E. (2020). Understanding work teams from a network perspective: A review and future research directions. *Journal of Management*, *46*(6), 1002–1028. https://doi.org/10.1177/0149206320901573

Park, S., Mathieu, J., & Grosser, T. J. (2020). A network conceptualization of team conflict. *Academy of Management Review*, *45*(2), 352–375. https://doi.org/10.5465/amr.2016.0472

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. University of Texas at Austin.

Ployhart, R. E., & Moliterno, T. P. (2011). Emergence of the human capital resource: A multilevel model. *Academy of Management Review*, *36*(1), 127–150. https://doi.org/10.5465/amr.2009.0318

Pollak, A. M., Elling, B., & Smith, M. (2012). *Nancy Caroline's Emergency Care in the Streets* (Vol. 2). Jones and Bartlett Learning.

Pratt, M. G., Kaplan, S., & Whittington, R. (2020). Editorial essay: The tumult over transparency: Decoupling transparency from replication in establishing trustworthy qualitative research. *Administrative Science Quarterly*, *65*(1), 1–19. https://doi.org/10.1177/0001839219887663

Reed, N., Metzger, Y., Kolbe, M., Zobel, S., & Boos, M. (2018). Unitizing verbal interaction data for coding. In E. Brauner, M. Boos, & M. Kolbe (Eds.), *The cambridge handbook of group interaction analysis* (pp. 208–226). Cambridge University Press. https://doi.org/10.1017/9781316286302.012

Rusko, M., & Finke, M. (2016). *Using speech analysis in voice communication: A new approach to improve air traffic management security* [Paper presentation] The 2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) (pp. 181–186), Wroclaw, Poland.

Schneider, B., González-Romá, V., Ostroff, C., & West, M. A. (2017). Organizational climate and culture: Reflections on the history of the constructs in the Journal of Applied Psychology. *Journal of Applied Psychology*, *102*(3), 468–482. https://doi.org/10.1037/apl0000090

Seelandt, S. (2018). Quality Control: Assessing reliability and validity. In E. Brauner, M. Boos, & M. Kolbe (Eds.), *The cambridge handbook of group interaction analysis* (pp. 227–244). Cambridge University Press. https://doi.org/10.1017/9781316286302.013

Sewell, D. K., & Chen, Y. (2016). Latent space models for dynamic networks with weighted edges. *Social Networks*, *44*, 105–116. https://doi.org/10.1016/j.socnet.2015.07.005

Shipp, A. J., & Cole, M. S. (2015). Time in individual-level organizational studies: What is it, how is it used, and why isn't it exploited more often? *Annual Review of Organizational Psychology and Organizational Behavior*, *2*(1), 237–260. https://doi.org/10.1146/annurev-orgpsych-032414-111245

Short, J. C., Broberg, J. C., Cogliser, C. C., & Brigham, K. H. (2010). Construct validation using computer-aided text analysis (CATA) an illustration using entrepreneurial orientation. *Organizational Research Methods*, *13*(2), 320–347. https://doi.org/10.1177/1094428109335949

Short, J. C., McKenny, A. F., & Reid, S. W. (2018). More than words? Computer-aided text analysis in organizational behavior and psychology research. *Annual Review of Organizational Psychology and Organizational Behavior*, *5*, 415–435. https://doi.org/10.1146/annurev-orgpsych-032117-104622

Slaughter, A. J., & Koehly, L. M. (2016). Multilevel models for social networks: Hierarchical Bayesian approaches to exponential random graph modeling. *Social Networks*, *44*, 334–345. https://doi.org/10.1016/j.socnet.2015.11.002

Smith-Jentsch, K. A. (2007). The impact of making targeted dimensions transparent on relations with typical performance predictors. *Human Performance*, *20*(3), 187–203. https://doi.org/10.1080/08959280701332992

Steffens, N. K., & Haslam, S. A. (2013). Power through 'us': Leaders' use of we-referencing language predicts election victory. *PLOS ONE*, *8*(10), Article e77952. https://doi.org/10.1371/journal.pone.0077952

Stephanian, D., Sawyer, T., Reid, J., Stone, K., Roberts, J., Thompson, D., & Pendergrass, T. (2015). Synchronous mobile audio-visual recording technology (SMART) cart for healthcare simulation debriefing. *Simulation & Gaming*, *46*(6), 857–867. https://doi.org/10.1177/1046878115617711

Sundstrom, E., De Meuse, K. P., & Futrell, D. (1990). Work teams: Applications and effectiveness. *American Psychologist*, *45*(2), 120–133. https://doi.org/10.1037/0003-066X.45.2.120

Tannenbaum, S. I., & Cerasoli, C. P. (2013). Do team and individual debriefs enhance performance? A meta-analysis. *Human Factors*, *55*(1), 231–245. https://doi.org/10.1177/0018720812448394

Waller, M. J., Okhuysen, G. A., & Saghafian, M. (2016). Conceptualizing emergent states: A strategy to advance the study of group dynamics. *The Academy of Management Annals*, *10*(1), 561–598. https://doi.org/10.5465/19416520.2016.1120958

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, *9*(1), 1–26. https://doi.org/10.1177/014662168500900101

Wolf, L. E., Patel, M. J., Williams Tarver, B. A., Austin, J. L., Dame, L. A., & Beskow, L. M. (2015). Certificates of confidentiality: Protecting human subject research data in law and practice. *The Journal of Law, Medicine & Ethics*, *43*(3), 594–609.

Zappa, P., & Robins, G. (2016). Organizational learning across multi-level networks. *Social Networks*, *44*, 295–306. https://doi.org/10.1016/j.socnet.2015.03.003

Received October 10, 2019
Revision received September 24, 2020
Accepted October 5, 2020 ■