2014

The Development and Testing of a Measurement System to Assess Intensive Care Unit Team Performance

Aaron Dietz
University of Central Florida

THE DEVELOPMENT AND TESTING OF A MEASUREMENT SYSTEM TO ASSESS INTENSIVE CARE UNIT TEAM PERFORMANCE

by

AARON S. DIETZ
B.A. University of Washington, 2007
M.A. University of Central Florida, 2013

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Fall Term
2014

Major Professor: Eduardo Salas

ABSTRACT

Teamwork is essential for ensuring the quality and safety of healthcare delivery in the intensive care unit (ICU). Complex procedures are conducted with a diverse team of clinicians with unique roles and responsibilities. Information about care plans and goals must also be developed, communicated, and coordinated across multiple disciplines and transferred effectively between shifts and personnel. The intricacies of routine care are compounded during emergency events, which require ICU teams to adapt to rapidly changing patient conditions while facing intense time pressure and conditional stress. Realities such as these emphasize the need for teamwork skills in the ICU.

The measurement of teamwork serves a number of different purposes, including routine assessment, directing feedback, and evaluating the impact of improvement initiatives. Yet no behavioral marker system exists in critical care for quantifying teamwork across multiple task types. This study contributes to the state of science and practice in critical care by taking a (1) theory-driven, (2) context-driven, and (3) psychometrically-driven approach to the development of a teamwork measure. The development of the marker system for the current study considered the state of science and practice surrounding teamwork in critical care, the application of behavioral marker systems across the healthcare community, and interviews with front line clinicians. The ICU behavioral marker system covers four core teamwork dimensions especially relevant to critical care teams: *Communication, Leadership, Backup and Supportive Behavior,* and *Team Decision Making*, with each dimension subsuming other relevant subdimensions.

This study provided an initial assessment of the reliability and validity of the marker system by focusing on a subset of teamwork competencies relevant to subset of team tasks. Two raters scored the performance of 50 teams along six subdimensions during rounds (n=25) and

handoffs (n=25). In addition to calculating traditional forms of reliability evidence [intraclass correlations (ICCs) and percent agreement], this study modeled the systematic variance in ratings associated with raters, instances of teamwork, subdimensions, and tasks by applying generalizability (G) theory. G theory was also employed to provide evidence that the marker system adequately distinguishes teamwork competencies targeted for measurement.

The marker system differentiated teamwork subdimensions when the data for rounds and handoffs were combined and when the data were examined separately by task (G coefficient greater than 0.80). Additionally, variance associated with instances of teamwork, subdimensions, and their interaction constituted the greatest proportion of variance in scores while variance associated with rater and task effects were minimal. That said, there remained a large percentage of residual error across analyses. Single measures ICCs were fair to good when the data for rounds and handoffs were combined depending on the competency assessed (0.52 to 0.74). The ICCs ranged from fair to good when only examining handoffs (0.47 to 0.69) and fair to excellent when only considering rounds (0.53 to 0.79). Average measures ICCs were always greater than single measures for each analysis, ranging from good to excellent (overall: 0.69 to 0.85, handoffs: 0.64 to 0.81, rounds: 0.70 to 0.89). In general, the percent of overall agreement was substandard, ranging from 0.44 to 0.80 across each task analysis. The percentage of scores within a single point, however, was nearly perfect, ranging from 0.80 to 1.00 for rounds and handoffs, handoffs, and rounds.

The confluence of evidence supported the expectation that the marker system differentiates among teamwork subdmensions. Yet different reliability indices suggested varying levels of confidence in rater consistency depending on the teamwork competency that was measured. Because this study applied a psychometric approach, areas for future development and testing to redress these issues were identified. There also is a need to assess the viability of this tool in other research contexts to evaluate its generalizability in places with different norms and organizational policies as well as for different tasks that emphasize different teamwork skills. Further, it is important to increase the number of users able to make assessments through low-cost, easily accessible rater training and guidance materials. Particular emphasis should be given to areas where rater reliability was less than ideal. This would allow future researchers to evaluate team performance, provide developmental feedback, and determine the impact of future teamwork improvement initiatives.

ACKNOWLEDGMENTS

I could not have completed this dissertation without the support of my family, friends, colleagues, and mentors. Foremost, I would like to thank my wife Alyssa. Completing this project while living in different states would not have been possible without her continuous positivity, encouragement, and patience.

I am beyond grateful to my advisor Eduardo Salas for his mentorship throughout my graduate career. I would also like to thank my other committee members: Florian Jentsch, Valerie Sims, and Shawn Burke for their advice, expertise, and feedback throughout this process and my graduate training. Last, I would like to thank Michael Rosen for his mentorship, guidance, and, of course, for waking up in the early morning hours to collect data with me.

Portions of this dissertation draw from previously published work I have coauthored with colleagues and are acknowledged in the footnote of chapters from which this material appears: Dietz et al. (2014a) and Dietz et al. (2014b). Dietz et al. (2014a) is reprinted from the Journal of Critical Care, 29/6, Aaron S. Dietz, Peter J. Pronovost, Pedro Alejandro Mendez-Tellez, Rhonda Wyskiel, Jill A. Marsteller, David A. Thompson, and Michael A. Rosen (authors), "A systematic review of teamwork in the intensive care unit: What do we know about teamwork, team tasks, and improvement strategies?" pages 908-914, copyright (2014), with permission from Elsevier. Dietz et al. (2014b) is reproduced from BMJ Quality and Safety, "A systematic review of behavioural marker systems in healthcare: What do we know about their attributes, validity, and application?" Aaron S. Dietz, Peter J. Pronovost, Kari N. Benson, Pedro Alejandro Mendez-Tellez, Cynthia Dwyer, Rhonda Wyskiel, and Michael A. Rosen (authors), advanced online publication, copyright (2014), with permission from BMJ Publishing Group Ltd.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION[1]

Statement of Problem

Intensive care units (ICUs) first emerged in the 1950s as an organizational strategy to improve the efficiency of care given to patients with life-threatening conditions [Joint Commission Resources (JCR, 2004)]. These units were established to centralize specialized staff and supporting technologies so hospitals could provide their most vulnerable patient populations with immediate accesses to critical care services (JCR, 2004). Today, ICU admissions surpass four million patients each year in the United States, and it is estimated that 80% of Americans will have some contact with these facilities in their life; some as a patient and others by extension as a family member or close friend (JCR, 2004). Profoundly, one-fifth of ICU patients will likely experience an injury from their treatment(s) while hospitalized in the ICU (JCR, 2004).

Breakdowns in teamwork have been recognized as a prominent contributor to medical errors and incidents across clinical domains for over two decades (Gawande, Zinner, Studdert, & Brennan, 2003; Leape et al., 1991) and ICUs are not immune from teamwork failures (Pronovost et al., 2006). In a single-center study, 37% of error records (205 of 554) involved verbal communication between nurses and physicians (Donchin et al., 1995). Similarly, a multi-center analysis of 2,075 incidents reported by 23 ICUs over a one year period revealed that team factors

[1] Segments of this chapter include previously published material of the author: Dietz et al. (2014a) and Dietz et al. (2014b). Dietz et al. (2014a) is reprinted from the Journal of Critical Care, 29/6, Aaron S. Dietz, Peter J. Pronovost, Pedro Alejandro Mendez-Tellez, Rhonda Wyskiel, Jill A. Marsteller, David A. Thompson, and Michael A. Rosen (authors), “A systematic review of teamwork in the intensive care unit: What do we know about teamwork, team tasks, and improvement strategies?” pages 908-914, copyright (2014), with permission from Elsevier. Dietz et al. (2014b) is reproduced from BMJ Quality and Safety, “A systematic review of behavioural marker systems in healthcare: What do we know about their attributes, validity, and application?” Aaron S. Dietz, Peter J. Pronovost, Kari N. Benson, Pedro Alejandro Mendez-Tellez, Cynthia Dwyer, Rhonda Wyskiel, and Michael A. Rosen (authors), advanced online publication, copyright (2014), with permission from BMJ Publishing Group Ltd.

such as communication, team structure, and leadership contributed to 32% of all incidents (Pronovost et al., 2006). These examples illustrate the saliency of teamwork breakdowns in the ICU and their potential to jeopardize patient safety.

Patient safety researchers and practitioners alike have turned their attention toward identifying effective team processes and improvement strategies as mechanisms to help eliminate preventable harm and death (Buljac-Samardzic, Dekker-van Doorn, van Wijngaarden, & van Wijk, 2010; Burke, Salas, Wilson-Donnelly, & Priest, 2004; Rosen, Salas, Silvestri, Wu, & Lazzara, 2008; Schmutz & Manser, 2013). The validity of conclusions drawn from these efforts, and the extent to which teamwork improves and safety outcomes are realized as a result of interventions is contingent upon rigorous, psychometrically driven measurement practices (Rosen et al., 2008). The development and use of valid team performance measures is essential for providing accurate assessments, directing feedback, and determining the impact of quality improvement initiatives (Rosen et al., 2010).

Despite the importance of team performance and its measurement in healthcare, a number of methodological challenges are widely prevalent—no matter the clinical context. Examples include the emphasis on training individual competencies, variability in work tasks, and the heterogeneity of team composition and structure (Jeffcott & Mackenzie, 2008). Additionally, the use of inconsistent terminology to describe teamwork constructs pervades the healthcare literature (Baker, Gustafson, Beaubien, Salas, & Barach, 2005). There also is considerable variation in how teamwork is conceptualized and operationalized (Baker et al., 2005). In fact, the Agency for Healthcare Quality and Safety (AHRQ) concluded that "medical teamwork and team training research are not formally linked to medical team performance theory" (Baker et al., 2005, p.50). To redress these gaps, the AHRQ outlined critical research needs, including the need for: (1) a medical team performance model, (2) teamwork process and outcome measures, relative to medicine, and (3) more efficient practices for evaluating medical team training programs.

The ICU teamwork literature faces similar theoretical challenges. As highlighted in Chapter 2 (and Chapter 3 for the healthcare community at large), issues related to construct clarification and terminology use abound. Practically, this means the link between teamwork and safety and performance outcomes may be misrepresented or misleading. It also complicates the comparison of research findings across studies; just because a researcher labels a construct as communication in their research, does not necessarily mean the measurement of that construct is appropriate given the theoretical basis of what communication represents. Such classification issues may preclude quantitative comparisons across study findings.

Purpose of the Current Study

The purpose of this study is to develop and test the psychometric properties of a behavioral marker system to assess ICU team performance across multiple task types. In order to develop and evaluate the potential of the marker system for assessing ICU team performance, this study will be: (1) theory-driven, (2) context-driven, and (3) psychometrically-driven. First, this study will develop a theoretical framework of ICU team performance to guide measurement. As noted above, the importance of rooting measurement in theory cannot be understated (Baker & Salas, 1992). The marker system will also be context-driven, meaning guidance from subject matter experts will be leveraged to provide valid accounts of how competencies relevant to critical care uniquely manifest. Last, the reliability and validity of the marker system will be evaluated to outline the strengths and shortcomings of the tool to facilitate future development. Twenty unique marker systems have been reported in medical team research, yet the psychometric evidence supporting the validity of these systems is often dubious (see Chapter 3). Failing to account for sources of systematic variance in observed scores may mean that measurements are representing variance that is not attributable to the construct(s) of interest; invalidity is the product of variance resulting from systematic effects other than those due to targeted constructs (DeShon, 2002). The present study will therefore explore systematic effects in scoring attributions by applying generalizability (G) theory to model good and bad variance associated with the measurement procedure. Traditional reliability and agreement indices will also be employed (e.g., intraclass correlations and rater agreement). By relying on a psychometric approach, specific areas for improvement can be identified.

Overall, this study is expected to provide both theoretical and practical contributions to the scientific community. First, this study advances the science of teams in critical care by clearly explicating what team behaviors matter most for different types of team tasks. An understanding of these factors will be informed from a review of the ICU teamwork literature at large, critical incident interviews with ICU clinicians, and mapping these findings to theoretical underpinnings of teamwork in general (see Chapter 4). Second, this study will apply a psychometric approach to understand where and why the marker system should be improved for future use and development, a strategy which is not always used in healthcare team research (Chapter 3). Following improvements identified through this study, the culmination of these efforts will ultimately result in an evaluation tool that can be used in ICUs to assess teamwork, facilitate feedback, and determine the impact of future teamwork improvement initiatives.

Background

This section briefly reviews the science of teams, team performance measurement, and behavioral marker systems to provide structure for the themes presented in this study.

The Science of Teams

A robust multidisciplinary science of teams has explicated a broad set of factors related to team effectiveness (Kozlowski & Ilgen, 2006; Salas, Cooke, & Rosen, 2008). This literature, as well as medical team research at large, has been inundated with inconsistent terminology. Consequently, key terms used in this study are defined. A *team* refers to a set of two or more individuals with specific roles who work interpedently and adaptively toward a shared goal (Salas, Dickinson, Converse, & Tannenbaum, 1992). Behavior within teams can be classified in terms of *taskwork* (i.e., behaviors related to how individual team member's carry out their individual work) and *teamwork* (i.e., behaviors involved when team members interact; Baker & Salas, 1992). *Team performance* is the confluence of taskwork and teamwork activities (i.e., what the team actually does) and *team performance effectiveness* refers to whether team performance outcomes fulfill performance goals and expectations (Salas, Stagl, Burke, & Goodwin, 2007). In healthcare, the term *nontechnical skills* also has been used to describe both individual- and team-related behaviors that are not related to technical aspects of clinical practice (Gordon, Darbyshire, & Baker, 2012; Yule & Paterson-Brown, 2012).

Team performance is generally characterized in terms of inputs, mediators, and outcomes (IMO) (Kozlowski & Ilgen, 2006). This IMO framework has been adopted in healthcare as well (Reader, Flin, Mearns, & Cuthbertson, 2009). As shown in Figure 1, the influence of input variables such as team, task, and environmental characteristics on focal performance outcomes (e.g., patient outcomes and team outcomes) is dependent on the effectiveness of team processes. Team processes are the dynamic interactions of team members and can broadly be categorized as transition (i.e., preparing for or reflecting on the team's work), action (i.e., task execution), or interpersonal (i.e., managing personal relationships) in nature (LePine, Piccolo, Jackson, Mathieu, & Saul, 2008; Marks, Mathieu, & Zaccaro, 2001). This general conceptualization serves as the organizational framework for this study. Chapter 4 details how teamwork theory was mapped with context-specific features of critical care to inform the development of the ICU behavioral marker system.

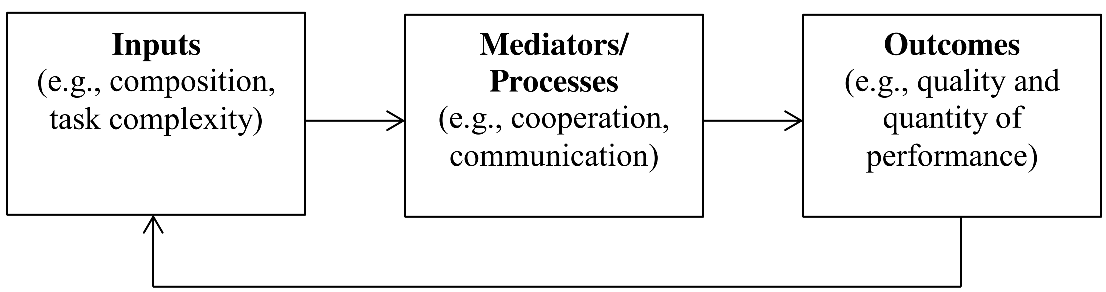

Figure 1. A generic IPO framework of ICU team performance.

Team Performance Measurement

Brannick and Prince (1997) eloquently stated "one can think of measurement as an investment in which one purchases information to inform a decision or some kind of action" (p. 5). Without measurement, researchers are unable to understand how well teams are performing, the impact of interventions (e.g., training), or direct developmental feedback. The authors go on to describe how factors related to the purpose of measurement, assessment context, attributes being measured, and quantifying those attributes influence the measurement of teamwork.

The measurement tool developed for the present study should serve a variety of purposes related to understanding and improving team functioning in critical care. The types of team tasks conducted in the ICU will shape what aspects of teamwork should be measured and when they should be measured (Chapter 4). For instance, responding to a code event will require team members to exhibit different competencies (e.g., supporting behaviors) than daily rounds (e.g., planning and establishing goals). The next step in measurement construction is to determine how to quantify relevant competencies. Communication, for example, is a latent construct central to team functioning. How is meaning ascribed to the quality of communication? To answer such a question, the present study will develop a behavioral marker system for understanding and quantifying teamwork competencies (see section below).

Once a measure has been developed, it is only useful to the extent that assessments are reliable and valid. Briefly, reliability concerns the consistency of measurements (e.g., overtime, between raters) while validity addresses the quality of inferences that can be made from measurement; if observed scores are the product of bias rather than manifestations of target constructs, the researcher is unable to draw valid conclusions about research findings. Additionally, multiple criteria should be leveraged to assess the psychometric properties of a measurement system; collecting multiple forms of reliability and validity evidence is the psychometric paragon for demonstrating a measurement tool is actually measuring what it intends to measure (e.g., Pedhazur & Schmelkin, 1991; Osterlind, 2010. These concepts are described in greater detail in Chapter 3 as they relate to psychometric evaluation.

Behavioral Marker Systems

In healthcare, team assessment strategies generally rely on perceptual surveys and/or observational techniques, each with inherent strengths and weaknesses (Rosen et al., 2012).

Observation is a widely used strategy for evaluating skills and competencies that drive team processes (Baker & Salas, 1997). Behavioral marker systems are observational approaches to team performance measurement that rely on trained raters to assess overt behaviors. These characteristics make marker systems uniquely suited to capture teamwork skills with enhanced objectivity (Flin & Martin, 2001). Marker systems also are competency-driven and afford a standardized lexicon to structure assessments and feedback because of their specificity (ANTS, 2012; NOTSS, 2012; Rosen et al., 2008).

Behavioral markers are "a prescribed set of behaviors indicative of some aspect of performance" (Flin & Martin, 2001, p. 96). The purpose of a marker system is to rate observable behaviors or events in order to make inferences about team skills and cognitions. For example, situational awareness (SA) is a cognitive construct that involves perception, comprehension, and anticipation (Endsley, 1995; Reader, Flin, Mearns, & Cuthbertson, 2011). Behavioral marker systems have evaluated SA by eliciting behaviors related to gathering information (e.g., cross-checking), recognizing and understanding (e.g., articulation of cues and their importance), and anticipation (e.g., actions taken to circumvent a problem; ANTS).

Manuscript Organization

This study seeks to evaluate the strengths and shortcomings of a behavioral marker system developed to assess ICU team performance. In order to accomplish the objectives outlined for this study (Table 1), this manuscript begins with two systematic literature reviews. Chapter 2 details the state of science surrounding ICU teamwork to aid in the identification of context-specific competencies that drive team performance and the conditions when they are relevant. Chapter 3 examines the application of behavioral marker systems in healthcare research to provide guidance on the content and structure of the marker system developed for this study and the norms for psychometric quality. These chapters also serve to highlight gaps in research the current study will redress. Next, Chapter 4 provides an overview of how the marker system was developed, illustrating how teamwork theory was mapped to marker system content. Chapter 5 describes the methodological approach for testing the reliability and validity of the behavioral marker system and Chapter 6 reports study findings. Chapter 7 describes the implications of study findings for future research and development.

Table 1. Overview of study approach and findings.

Chapter	Key Points
Chapter 1: Introduction	• The measurement of teamwork serves a number of different purposes, including routine assessment, directing feedback, and evaluating the impact of improvement initiatives. • No measurement systems have been developed specifically to assess behavioral indicators of ICU team performance. Therefore, the aim of the present study was to develop and evaluate the potential of a behavioral marker system for assessing ICU team performance.
Chapter 2: Teamwork in the ICU	• This review helped identify competencies particularly relevant to ICU teams and what circumstances they are important (i.e., *what to measure*). • Implications of this review for the current study are addressed.
Chapter 3: Behavioral Marker Systems in Healthcare	• This review outlined key features of existing marker systems to guide the development of the marker system developed for this study (e.g., content, structure, psychometric evidence; i.e., *how to measure*). • Implications of this review for the current study are addressed.
Chapter 4: Behavioral Marker System Development Approach	• A framework is presented that demonstrates how teamwork theory was mapped to context-specific examples of performance to generate the behavioral marker system.
Chapter 5 Methodology	• A multifaceted approach to test the reliability and validity of the marker system is presented, including studies to index interrater reliability and agreement, establish content validity (literature reviews and critical incident interviews), and construct validity (G theory). • The methodological approach of the G study is outlined and relevant predications are made.
Chapter 6: Results	• The results of the G study provided support that the marker system reliably differentiates teamwork competencies and that there was no systematic variance associated with how raters scored certain teams or teamwork competencies. • Overall, interrater reliability was good, but analyses along specific competencies and tasks revealed areas that warrant further consideration to improve the tool prior to implementing the marker system further.

Chapter	Key Points
Chapter 7: Discussion	• This study provided initial validity evidence that the marker system can have utility in differentiating among teamwork competencies. • Efforts should be taken to study the validity of the marker system in other ICU settings and for a more representative sample of ICU tasks. • Rater training and guidance materials should be developed to redress limitations associated with rater reliability and agreement in order to help advance further use of the ICU marker system.

CHAPTER 2: TEAMWORK IN THE ICU[2]

Chapter Overview

The purpose of this chapter is to provide insight into what teamwork dimensions are particularly relevant for ICU teams and when these dimensions are important. This chapter begins by presenting key research questions to identify these factors and the methods used to systematically review literature content. Briefly, a great deal of ICU team-based research has focused on examining teamwork behaviors both in relation to a specific task as well as outside a specific task. Communication was the most widely cited teamwork construct and transition cycles of teamwork were the most emphasized type of team task investigated. Yet conceptualizations and operationalizations of teamwork varied across studies. The chapter concludes by discussing the implications of these findings for the current study.

Introduction

Increased recognition that teamwork plays a central role in patient safety has resulted in a rapid expansion of research examining teamwork, team tasks, and interventions to foster teamwork in the ICU (Ohlinger et al., 2003; Sexton et al., 2011; Stockwell, Slonim, & Pollack, 2007; Thomas, Sexton, & Helmreich, 2003). Several systematic and unsystematic reviews of teamwork have been conducted in this setting (Baggs, Norton, Schmitt, & Sellers, 2004; Lin, Chaboyer, & Wallis, 2009; Reader, Flin, Lauche, & Cuthbertson, 2006; Reader, Flin, & Cuthbertson, 2007; Reader et al., 2009). Given the accelerated pace of this research since

[2] This chapter includes previously published material of the author: Dietz et al. (2014a). Dietz et al. (2014a) is reprinted from the Journal of Critical Care, 29/6, Aaron S. Dietz, Peter J. Pronovost, Pedro Alejandro Mendez-Tellez, Rhonda Wyskiel, Jill A. Marsteller, David A. Thompson, and Michael A. Rosen (authors), "A systematic review of teamwork in the intensive care unit: What do we know about teamwork, team tasks, and improvement strategies?" pages 908-914, copyright (2014), with permission from Elsevier.

previous reviews, the greater variety of research offers an opportunity to provide a more detailed analysis of the types of team constructs under investigation, the clinical tasks that depend on teamwork, and interventions to optimize teamwork. Specifically, this chapter seeks to answer four key questions about teamwork in ICUs to help establish the theoretical underpinnings of the behavioral marker system. First, what is known about the nature and prevalence of team related failures? Second, how have researchers conceptualized teamwork in ICUs? Third, where has teamwork been investigated in ICUs (i.e., what tasks or work contexts)? Fourth, what interventions have been used to improve teamwork (i.e., for what competencies and under what conditions)? Answering these questions will inform the development of a theoretical framework of ICU team performance as well as outline strengths and limitations of the current state of ICU team research.

Method

A Boolean key word search of PubMed was conducted in February 2013. Key word combinations focused on three areas: (1) teamwork, (2) the ICU, and (3), interventions (e.g., training, quality improvement initiatives). Figure 2 details the screening process which was designed to capture the full spectrum of articles related to teamwork in the ICU. The coding scheme used to capture content was iteratively revised to ensure extracted content was relevant and meaningful to the aims of this review. Key variables included information about team processes and emergent states, team tasks, team interventions, and study outcomes (see Appendix A and B). After the final stage of screening (Figure 2), 85 articles were retained for further coding.

Stage 1: Boolean Search of Pubmed February 22, 2013
Results: Results: 3023 articles

↓

Stage 2: Screening of Titles
Inclusion Criteria: Nominal relevance to teamwork, team tasks, and/or team interventions
Results: 714 articles

↓

Stage 3: Screening of Abstracts
Inclusion Criteria:
1. Context of manuscript is the ICU AND
2. Manuscript investigates teamwork OR
3. Manuscript investigates team tasks (e.g., handover, transfer, rounds) OR
4. Manuscript investigates interventions to improve team processes OR team performance episodes

Results: 296

↓

Stage 4: Screening of Abstracts and Full Articles
Inclusion Criteria:
1. Context of manuscript is the ICU AND
2. Context of manuscript is intradepartment (i.e., everything within and only within the ICU) AND
3. Manuscript reviews a primary data source/includes some sort of replicable methodology AND
4. Manuscripts describing metric developments must describe results beyond the psychometric properties of the rating system OR
5. Must be able to extract data from ICU when multiple unit types are discussed

Results: 85

Figure 2. Overview of methodological approach for ICU team research review.

Results

Prevalence of Team Related Failures

A deep understanding of the nature of a problem facilitates the development of effective assessments and what factors should be targeted for measurement. Only two studies directly investigated the prevalence or nature of team failures in the ICU. These studies converge and suggest that approximately one-third of errors and adverse incidents were linked to communication (Donchin et al., 1995; Pronovost et al., 2006), leadership, and team structure (Pronovost et al., 2006). Several other studies have demonstrated significant associations between the level of teamwork and ICU outcomes. Positive caregiver interaction among ICU clinicians was associated with shortened length of stay (Shortell et al., 1994). Better leadership, conflict resolution, and coordination were associated with lower incidents of periventricular/intraventricular hemorrhage or periventricular leukomalacia (PIVH/PVL; (Pollack & Koch, 2003). Positive perceptions of nurse-physician collaboration were associated with reduced likelihood of mortality and/or readmission (Baggs et al., 1999).

Conceptualizations of ICU Teamwork

Teamwork is a broad construct with varying definitions. For example, Salas et al. (2008) reported that over 130 models and frameworks of team performance have been presented in the scientific literature. Understanding how investigators have applied teamwork to the ICU environment can provide guidance on what aspects of teamwork matter most in the ICU as well as how those concepts can be translated into practical guidance for measurement development. Twenty-seven unique constructs were identified (Table 2). In some cases, unique teamwork constructs were collapsed into a single category because of similarity in focus (e.g., team climate and culture). Seventeen percent of articles (n=14) did not explore any teamwork construct. These studies often examined a teamwork activity (e.g., rounds) or a teamwork intervention (e.g., documentation tool) in relation to patient, individual, or unit/organization outcomes. Many studies investigated more than one aspect of teamwork. The most widely studied construct was communication (n=44; 52%), followed by leadership (n=17; 20%), collaboration (n=16; 19%), coordination (n=12; 14%), and team climate/culture (n=7; 8%). Team constructs also were described with varying levels of specificity. For instance, many studies investigated or described communication as a unidimensional construct (Newkirk, Pamplin, Kuwamoto, Allen, & Chung, 2012; Sluiter et al., 2005) while other studies explored facets of communication such as closed-loop communication (Figueroa, Sepanski, Goldberg, & Shah, 2013) and the openness/quality of communication (Jukkala, James, Autrey, Azuero, & Miltner, 2012).

There also was a great deal of overlap in how team constructs were conceptualized. For instance, Boyle and Kochinda (2004) described collaborative communication to be the product of factors such as leadership, communication, coordination, problem-solving and conflict management, and team culture. Thomas et al. (2003) rated aspects of assertiveness, collaboration, cooperation, support, coordination, and conflict resolution to assess teamwork climate. Last, Miller (2001) measured leadership, the openness, satisfaction, and timeliness of communication, and problem-solving to gauge collaborative interaction.

Table 2. Summary of teamwork constructs, team tasks, and training interventions.

Research Question	**Results**
What outcomes are being investigated in ICU team research? (n=85)	• Team (44; 52%) • Task (43; 51%) • Patient (24; 28%) • Individual (20; 24%) • Unit/Organization (13; 15%)
How has teamwork been conceptualized and operationalized in the ICU? (n=85)	• Communication (44; 52%) • Leadership (17; 20%) • Collaboration (16; 19%) • Coordination (12; 14%) • Team Climate/Culture (7; 8%) • Information Exchange (3; 4%) • Conflict Management (3; 4%) • Cohesion (2; 2%) • SA/Team SA (2; 2%) • Shared Mental Model (2; 2%) • Assertion (1; 1%) • Caregiver Interaction (1; 1%) • Cooperation (1; 1%) • Decision-making Inclusion (1; 1%)

Research Question	Results
	• Empowerment (1; 1%) • Joint Sense-Making (1; 1%) • Mutual Performance Monitoring (1; 1%) • Mutual Respect (1; 1%) • Mutual Support/Assertion (1; 1%) • Shared Goal Agreement (1; 1%) • Shared Problem Solving (1; 1%) • Situation Monitoring (1; 1%) • Team Commitment (1; 1%) • Team Satisfaction (1; 1%) • Trust (1; 1%) • Verbalizing situational information (1; 1%) • Not Specified (14; 17%)
Where has teamwork been investigated in ICU? (n=85)	*Research Context* • Multiple ICUs (31; 37%) • General ICU (15; 18%) • Pediatric ICU (10; 12%) • Medical ICU (5; 6%) • Medical-Surgical ICU (5; 6%) • Surgical ICU (4; 5%) • Neurovascular ICU (3; 4%) • Neonatal ICU (2; 2%) • Pediatric Cardiac ICU (2; 2%) • Cardiothoracic ICU (1; 1%) • Medical-Surgical Pediatric ICU (1; 1%) • Neuro-ICU (1; 1%) • Neuroscience ICU (1; 1%) • Newborn ICU (1; 1%) • Neurosurgical ICU (1; 1%) • Trauma ICU (1; 1%) • Not Applicable (1; 1%) *Team Tasks* • Rounds (33; 39%) • Clinical (17; 20%) • Handoff (17; 20%) • Transfer (2; 2%) • Huddle/Debrief (1; 1%) • Multidisciplinary Meetings (1; 1%)
What interventions have been used to improve teamwork in the ICU? (n=36)	• Standardized Patient Status Tool (15; 42%) • Training (8; 22%) • Rounds/Change of Rounding Process (7; 19%) • Specialized Staffing (3; 8%) • Comprehensive Unit-Based Safety Program (2; 6%) • Robotic Tele-Presence (1; 3%) • Safety Attitude Questionnaire Action Plan (1; 3%) • Multidisciplinary Work Shift Evaluations (1; 3%) • Collaborative Communication Intervention (1; 3%) • Wireless Email (1; 3%)

Research Question	Results
Where has evidence of intervention effectiveness been demonstrated? (n=36)	• Team outcomes (18; 50%) • Task outcomes (21; 58%) • Patient outcomes (14; 39%) • Individual outcomes (9; 25%) • Unit/Organization outcomes (4; 11%)

Research Settings and Tasks

Reviewing the types of tasks or settings where teamwork has been investigated provides insight into where teamwork may be most important within an ICU, or what aspects of teamwork are most important under certain conditions.

Research Context

The context of research varied greatly across studies, indicating the relevance of teamwork across a wide variety of ICU-types (Table 2). Thirty-seven percent of studies (n=31) involved more than one ICU, and the majority of single ICU studies (n=38; 45%) had a unique clinical focus (e.g., pediatric, medical, surgical). One study examined attributes of leadership and leadership training at a workshop for pediatric intensivists (Stockwell, Pollack, Turenne, & Slonim, 2005).

Team Tasks

Team tasks investigated in ICU team research are summarized in Table 2. Rounds were the most common task described in articles (n=33; 39%) and can be characterized as a transition cycle of teamwork. Rounds typically involve a 20-25 minute discussion of each patient in which the clinical team prioritizes a daily plan of care (Pronovost et al., 2003). Clinicians can spend as much as 75% of their time engaged in communication events during rounds (Alvarez & Coiera, 2005). Rounds are a critical team task because they provide a forum in which the entire care team can communicate, yet are not necessarily the panacea for the formation of shared expectations for patient treatment (Custer et al., 2012). The effectiveness of rounds may be impeded by communication interruptions (Alvarez & Coiera, 2005) or the focus of conversation (e.g., provider-focused vs. goal-focused; Pronovost et al., 2003). Space constraints, time pressure, and inefficient access to patient information can further complicate the effectiveness of rounds (Ho, Xiao, Vaidya, & Hu, 2007).

Like rounds, handoffs can also be characterized as a transition cycle of teamwork and were described in 20% of reviewed articles (n=17). Handoffs primarily involve the coordination of patient care (Douglas et al., 2013). During one type of handoff, clinicians from an outgoing shift brief oncoming clinicians on a patients' status (Collins et al., 2012). The exchange of patient information is both complex and central to patient safety (Collins et al., 2012). Pronovost et al. (2006) found that 12% of incidents reported by 23 ICUs over a one year period resulted from breakdowns in verbal or written communication during handoffs. Ilan et al. (2012) observed that physicians spend about 3 minutes discussing each patient during end-of-week handoffs and that the appropriate use of standardized communication tools was inconsistent. Further, explicit recommendations were omitted in 60% of observations. Finally, Collins et al. (2012) reported that handoffs were generally a discipline-specific activity (e.g., nurse-nurse, physician-physician), limiting information sharing across roles.

For the purpose of this review, clinical tasks are broadly defined as specific taskwork activities such as cardiac arrest management (Figueroa et al., 2013) or when 'routine care' (Pronovost et al., 2006) was mentioned in an article. Clinical tasks were described in 20% of articles (n=17) and represent action phases of teamwork. In the ICU, work is often conducted at an accelerated pace to respond to changing patient conditions (Douglas et al., 2013). Task diversity magnifies the importance of communication and coordination (Shortell et al., 1994). For example, the perceived effectiveness of caregiver interaction was associated with better perceptions of technical care and increased ability to meet patient-family needs (Shortell et al., 1994). Communicating priorities and appropriate task delegation by leadership are also central to team performance (Reader et al., 2007).

Interventions to Improve Teamwork

Thirty-six articles described interventions to improve teamwork. As summarized in Table 2, many of these studies involved more than one intervention (e.g., multiple patient tools) and most were developed primarily to improve teamwork (n=22; 61%). The majority of interventions identified were standardized protocols (e.g., daily checklist, patient charts; n=15; 42%), implementation of daily rounds or modification to the rounding structure/process (n=7; 19%), and training (n=8; 22%).

Standardized protocols are typically applied to augment the rounding or handover process (Abraham, Kannampallil, & Patel, 2012; Narasimhan, Eisen, Mahoney, Acerra, & Rosen, 2006; Newkirk et al., 2012). Pronovost et al. (2003) developed a daily goal sheet as a communication tool to increase clinician understanding of patient care objectives for that day. Daily goals help to make goals explicit and reduce ambiguity among team members, especially when a team member reads back the patients goals. Prior to the intervention, daily patient goals were understood by less than 10% of residents and nurses. Following the intervention, daily patient goals were understood by more than 95% of nurses and residents. Daily goal sheets have been applied in a number of ICUs, given their effectiveness as a mechanism to improve the communication of daily care plans (Narasimhan et al., 2006; Phipps & Thomas, 2007; Rehder et al., 2012), but ensuring clinician compliance is a key challenge for realizing the benefit of these tools (Newkirk et al., 2012).

Rounds were described earlier as an important team task in which care plans are formally discussed and prioritized. Rounds led by an ICU physician have been associated with shorter hospital stays, reduced hospital costs, and fewer postoperative complications (Dimick, Pronovost, Heitmiller, & Lipsett, 2001). An explicit approach to rounds increased confidence among clinicians that a long-term care plan was in place for patients as well as their overall satisfaction with rounding processes (Dodek & Raboud, 2003). The implementation of multidisciplinary rounds also contributed to decreased incidents of adverse clinical outcomes (e.g., ventilator associated pneumonia, bloodstream infections, and urinary tract infections; Jain, Miller, Belt, King, & Berwick, 2006; Johnson et al., 2009).

All training interventions were designed specifically to improve teamwork and seven of eight training articles described interventions to improve teamwork skills during clinical tasks. There was not enough information to determine a specific task for one training article (Boyle & Kochinda, 2004). Simulation-based training was applied in five studies and in each case, high-fidelity simulators were used (Allan et al., 2010; Figueroa et al., 2013; Meurling, Hedman, Sandahl, Fellander-Tsai, & Wallin, 2013; Nunnink, Welsh, Abbey, & Buschel, 2009; Pascual et al., 2011). All studies reported improved team outcomes following team training. For example, Mayer et al. (2011) investigated team performance before and after a classroom-based course emphasizing the TeamSTEPPS® curriculum. Core competency areas such as communication, leadership, situation monitoring, and mutual support/assertion were significantly improved one-month following the intervention. Improvement was not significantly maintained for all of the competency areas 12-months after team training, suggesting the need for recurrent team training or other interventions to sustain program success. Allan et al. (2010) applied Crew Resource Management (CRM) principles to improve teamwork skills during resuscitation events. Following training, participants were more confident in their ability to lead future resuscitations and indicated they were more likely to speak up if they believed the resuscitation was not being managed effectively.

In sum, effective team interventions in the ICU include implementing rounds, standardizing the rounding process with daily goals, and enhancing teamwork skills through team training. No study evaluated the synergistic impact of all of three of these interventions. Evidence of intervention effectiveness has been demonstrated with respect to team factors (n=18; 50%; e.g., improved perception of communication after training; Meurling et al., 2013) patient factors (n=14; 39%; e.g., rates of ventilator associated pneumonia; Stone et al., 2011), task factors (n=21; 58%; e.g., perceived accuracy with a new sign-out document; Palma, Sharek, & Longhurst, 2011), individual factors (n=9; 25%; e.g., job satisfaction; Boyle & Kochinda, 2004), and unit/organizational factors (n=4; 11%; e.g., safety climate; Vigorito, McNicoll, Adams, & Sexton, 2011).

Summary

This chapter has enumerated which teamwork constructs have received the most attention in ICU team research as well as the team tasks and interventions investigated to guide the development of the behavioral marker system. Table 3 integrates key findings of this review by addressing which aspects of teamwork have been investigated in different task settings and targeted by which improvement methods. Team interventions are presented along the horizontal row and team tasks are presented along the vertical column. Within each cell are team constructs that were associated with tasks and/or interventions. The number of times a teamwork construct was specified in an article is noted in parentheses, with the exception of instances when a construct was only referenced once.

As demonstrated by Table 3, a great deal of research has focused on examining teamwork behaviors both in relation to a specific task as well as outside of a specific task. This is consistent with calls for measurement and training to focus on team components that are both general and specific to types of teams as well as general and specific to types of team tasks (Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995; Thomas, 2011). Communication was described earlier in this review as the most prominent teamwork construct identified. This it is not surprising because communication skills (e.g., clarity, completeness; Smith-Jentsch, Cannon-Bowers, Tannenbaum, & Salas, 2008) are needed during both transition and action cycles of teamwork. Transition cycles of teamwork were the most common team task investigated and, in turn, interventions to improve performance on such tasks were emphasized in the literature. Team training interventions targeted a variety of competencies needed to preform specific clinical tasks, while structured protocols were widely employed for improving the efficacy of rounds and handoffs. Interpersonal processes transpire during both transition and action cycles of teamwork (Marks et al., 2001). For example, Studdert et al. (2003) noted that the source of intrateam conflict typically centered on care plan discrepancies (55%), communication deficiencies with other team members (12%), not including all team members in decision making (9%), and poor coordination (7%). Team leadership can also have a profound influence on interpersonal processes; team leaders set the tone for positive teamwork by establishing norms to promote effective team member interactions, being accessible and encouraging, and providing constructive feedback (Reader, Flin, & Cuthbertson, 2011).

Implications

This chapter also illuminates areas that must be addressed before a behavioral marker system of ICU teamwork can be developed. First, surprisingly few studies examined the nature of teamwork breakdowns. Additional qualitative research (e.g., cognitive task analyses) is needed to better understand teamwork facilitators and barriers. Such an understanding can inform areas for measurement tools to target for assessment in order to provide clinical teams feedback on current levels of teamwork. Critical incident interviews with a diverse representation of clinicians were conducted to this end (Chapter 4).

Second, conceptualizations of teamwork varied in the ICU teamwork literature. This finding is consistent with the broader teamwork literature in healthcare (Baker et al., 2005) and represents low hanging fruit that can yield significant dividends. For example, communication has been conceptualized in the ICU team literature as (1) a unidimensional construct, (2) a multidimensional construct, and (3) an attribute of other constructs. With respect to developing a measurement tool, this finding signifies the importance of explicitly outlining teamwork competencies in relation to existing definitions of teamwork dimensions. A key challenge, however, is that teamwork constructs are not orthogonal (LePine et al., 2008). Communication, coordination, and cooperation skills are all likely to contribute to the effective exchange of patient information. For example, teamwork climate as described by Thomas et al. (2003) is a manifestation of several dimensions. A critical research need for the present study is to clearly delineate what attributes are being measured and report findings in relationship to a clearly defined theoretical and operational definition of the construct(s) under investigation. Such construct clarification will allow for more meaningful interpretations of study findings and provide a foundation on which to base future quantitative reviews of teamwork within the ICU.

At this point, it is important to note the work of Reader and colleagues (2009) towards redressing this gap. The authors presented a theoretical framework of ICU team performance, where factors related to team input variables (team, task, and leader characteristics) are mediated by teamwork processes (communication, leadership, coordination, and decision-making) to produce team outputs (team outcomes and patient outcomes). This chapter builds on this work by providing a more detailed analysis of the types of team constructs under investigation in relation to team tasks and interventions to guide the development of the behavioral marker system.

Third, teamwork has been investigated across a wide range of ICU types and a strong emphasis has been placed on transition cycles of teamwork (i.e., rounds and handoffs). Further, a great deal of research has focused on examining teamwork behaviors both in relation to a specific task as well as outside of a specific task. This is consistent with calls for both general and task specific interventions to improve teamwork (Cannon-Bowers et al., 1995; Thomas, 2011) and reinforces the conception of patient care in the ICU as a complex team endeavor. Depending on the type of team task, there may be variability in team composition, the degree of interdependence required, and the pace at which tasks must be completed. This reality requires team members to develop competencies that are not only specific to a particular task or team (e.g., implicit coordination, shared mental models), but also competencies that are transportable and can be generalized to different teams and different tasks (e.g., assertiveness, backup behavior; Cannon-Bowers et al., 1995). Future research would benefit from explicitly defining the functional characteristics of team tasks that are investigated, the competencies required for task execution, and whether the competencies are specific or generic to ICU teams and tasks. A theoretical framework specifying aspects of teamwork that matter for ICU team performance and when they matter is presented in Chapter 4. This framework is important to convey whether manifestations of teamwork constructs vary by task and will help future researchers interpret the extent to which study results generalize to new ICU team contexts.

Conclusions

This chapter provided insight into what teamwork competencies are most relevant to ICUs and the tasks when they are most relevant. Additionally, this chapter identified research areas that must be addressed before a behavioral marker system of ICU teamwork can be developed, including: (1) additional qualitative research to understand drivers and barriers to ICU team performance, (2) explicitly defining teamwork constructs, and (3) explicating a theoretical framework of ICU team performance that specifies what dimensions of teamwork matter most and when they matter (see Chapter 4).

Table 3.Team Construct X Task X Intervention Matrix.

Interven-tion ⩾ Task ⩒	**Training (n=8)**	**Rounds/ Change of Round-ing Process (n=7)**	**Robotic Tele-presence (n=1)**	**Standard-ized Patient Status Tool (n=15)**	**Safety Attitude Question-naire Action Plan (n =1)**	**Staffing (n=3)**	**CUSP (n=2)**	**Collaborative Comm-unication Intervention (n=1)**	**Wire-less Email (n=1)**	**Multi-disciplinary Work Shift Evaluations (n =1)**	**Not Specified/ Not Applic-able (n =49)**
Clinical (n = 17)	1. Comm-unication (4) 2. Collabor-ation 3. Team Climate 4. Mutual Respect 5. Empower-ment 6. Leader-ship (4) 7. Situation Monit-oring 8. Mutual Support/ Assertion 9. Situation-al Aware-ness 10. Assertion (2) 11. Coord-ination 12. Mutual perform-ance			1. Comm-unication (1)		1. Comm-unication 2. Not specified					1. Comm-unication (4) 2. Leadership (2) 3. Coord-ination 4. Decision-making Inclusion 5. Collab-oration (2) 6. Leadership 7. Cohesion 8. Not specified

Interven-tion → Task ↓	Training (n=8)	Rounds/ Change of Round-ing Process (n=7)	Robotic Tele-presence (n=1)	Standard-ized Patient Status Tool (n=15)	Safety Attitude Question-naire Action Plan (n =1)	Staffing (n=3)	CUSP (n=2)	Collaborative Comm-unication Intervention (n=1)	Wire-less Email (n=1)	Multi-disciplinary Work Shift Evaluations (n =1)	Not Specified/ Not Applic-able (n =49)
	monitor-ing 13. Verbal-izing situation-nal inform-ation 14. Not specified										
Handoff (n =17)				1. Comm-unication (2) 2. Not specified (2)							1. Coord-ination (6) 2. Comm-unication (10) 3.Info-rmation exchange (2) 4. Collab-oration (2) 5. Shared mental model (2)

Interven-tion → Task ↓	Training (n=8)	Rounds/ Change of Round-ing Process (n=7)	Robotic Tele-presence (n=1)	Standard-ized Patient Status Tool (n=15)	Safety Attitude Question-naire Action Plan (n =1)	Staffing (n=3)	CUSP (n=2)	Collaborative Comm-unication Intervention (n=1)	Wire-less Email (n=1)	Multi-disciplinary Work Shift Evaluations (n =1)	Not Specified/ Not Applic-able (n =49)
Rounds (n = 33)		1. Comm-unication (4) 2. Cohesion 2. Not Specified (2)	1. Team Satisf-action	1. Comm-unication (8) 2. Team Culture (2) 3. Shared Goal Agree-ment 4. Collabor-ation 5. Not Specified (2)		1. Not specified					1. Coord-ination (3) 2. Comm-unication (11) 3. Leadership (4) 4. Team SA 5. Info-rmation exchange (2) 5.Shared mental model (2) 6. Collab-oration (3) 7. Cohesion 8. Joint sense-making 9. Not specified (3)
Huddle/ Debrief (n =1)											1. N/A (study dependent variable)
Transfer (n =2)											1. Collab-oration (2)
Multi-disciplin-ary Meetings											1. Collab-oration

Interven- tion → Task ↓	Training (n=8)	Rounds/ Change of Round- ing Process (n=7)	Robotic Tele- presence (n=1)	Standard- ized Patient Status Tool (n=15)	Safety Attitude Question- naire Action Plan (n =1)	Staffing (n=3)	CUSP (n=2)	Collaborative Comm- unication Intervention (n=1)	Wire- less Email (n=1)	Multi- disciplinary Work Shift Evaluations (n =1)	Not Specified/ Not Applic- able (n =49)
(n=1)											
Not Specified/ Not Applicable (n = 25)	1. Comm- unication	1. Not Specified			1. Team Climate		1. Team Climate (2)	1. Comm- unication	1. Not specified	1. Comm- unication	1. Collab- oration (8) 2. Comm- unication (6) 3. Leadership (7) 4. Coord- ination (4) 5. Team Commit- ment 6. Team Climate (2) 7. Conflict manage- ment (3) 8. Caregiver Interaction 9. Shared problem solving 10. Coop- eration 11. Trust 12. Not specified (2)

CHAPTER 3: BEHAVIORAL MARKER SYSTEMS IN HEALTHCARE[3]

Chapter Overview

With an understanding of the state of ICU team research in mind, this chapter systematically reviews of the use of behavioral marker systems in medical team research to provide guidance on how the marker system developed for the present study should be structured as well as the evidence needed to establish psychometric quality. Four key research questions will be addressed: (1) what are the attributes of the behavioral marker systems used in healthcare, (2) what evidence of reliability and validity exist, (3) what skills and expertise are required for their use, and (4) how have behavioral marker systems been applied to investigate the relationship between teamwork and other constructs in healthcare?

Method

A Boolean search was conducted using PubMed in February 2013 to identify articles relating to the following components: (1) health professionals/healthcare, (2) teamwork/ nontechnical skills, and (3) behavioral assessment/observation. Figure 3 summarizes the screening process. A coding scheme was iteratively developed to systematically capture article content germane to the objectives of this review, including: attributes of the marker system (i.e., behaviors, techniques, targets of measurement), psychometric properties of the marker system, and application of the marker system in healthcare research.

---

[3] This chapter includes previously published material of the author: Dietz et al. (2014b). Dietz et al. (2014b) is reproduced from BMJ Quality and Safety, "A systematic review of behavioural marker systems in healthcare: What do we know about their attributes, validity, and application?" Aaron S. Dietz, Peter J. Pronovost, Kari N. Benson, Pedro Alejandro Mendez-Tellez, Cynthia Dwyer, Rhonda Wyskiel, and Michael A. Rosen (authors), advanced online publication, copyright (2014), with permission from BMJ Publishing Group Ltd.

Results

Thirty-seven articles describing 20 unique marker systems met the inclusion criteria (Figure 3). The most widely employed marker system was the Observational Teamwork Assessment for Surgery (OTAS) evaluation tool (n=10; 27%). No other marker system was described more than three times. The primary purpose of 60% of articles (n=22) was to report development or validation efforts of the marker system. Eighty-nine percent of articles were quantitative (n=33) compared to just 11% that were qualitative (n=4). Sixty percent of articles specified that raters received some sort of training prior to behavioral assessment (n=22). Rater training was not specified in 30% of quantitative articles (n=11).

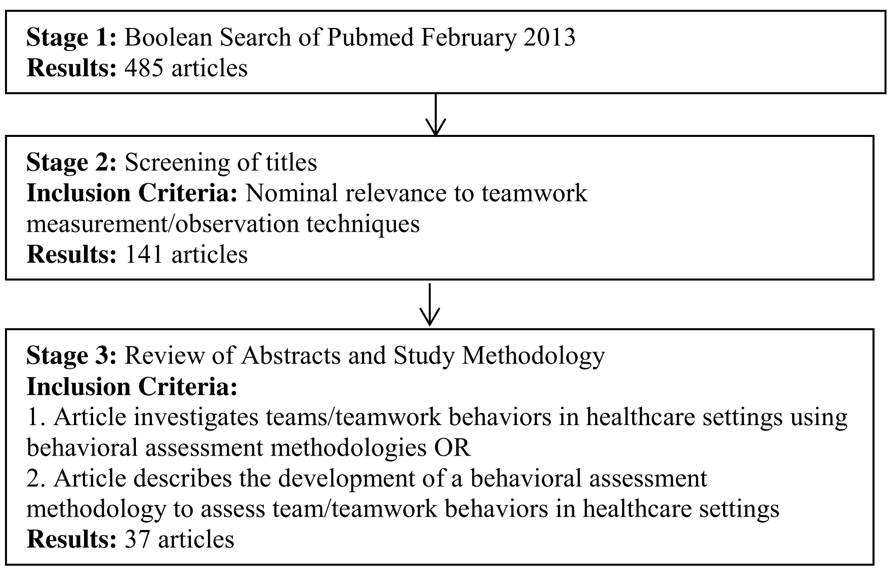

Figure 3. Methodological approach for marker review.

Attributes of Behavioral Marker Systems

This question addresses what behaviors are being investigated, why they are being investigated, and what techniques are used for assessment.

Content of Measurement

The systems reviewed applied a variety of classification structures varying in their level of specificity or granularity. Of the 20 identified measurement systems, six utilized a hierarchical structure to cluster behaviors. For example, the Non-Technical Skills for Surgeons (NOTSS) system obtains ratings for four categories of behavior (situation awareness, decision-making, communication and teamwork, and leadership) each with three elements that constitute a taxonomy of nontechnical skills for surgeons (NOTSS, 2012). Each NOTSS element is paired with positive and negative examples of behaviors to guide assessment. Other systems, such as OTAS, do not categorize behaviors with subdimensions (OTAS, 2011). The Just-In-Time Pediatric Airway Provider Performance Scale (JIT-PAPPS; Nishisaki et al., 2011) assesses decision-making as a unidimensional construct while the Anaesthetists' Non-Technical Skills (ANTS) system assesses decision-making as the product of (1) identifying options, (2) balancing and selecting options, and (3) re-evaluating (ANTS, 2012; Fletcher et al., 2003).

In order to examine what behaviors were targeted for measurement, behaviors (both categories and elements) from each marker system were amalgamated. One-hundred and four unique behaviors were identified after exact duplicates were removed. Next, duplicates with nominal relevance were removed to account for redundancies in terminology that are ostensibly describing the same attribute (e.g., coordination, coordinating activities with team members, coordinating with others). Seventy-nine unique constructs were retained following this qualitative data reduction. There were other instances where behaviors were paired with discrete constructs [e.g., leadership and team coordination (Frengley et al., 2011), teamwork and cooperation (Flowerdew, Brown, Vincent, & Woloshynowych, 2012)].

Context of Measurement

Fifteen of 20 marker systems were developed for a specific clinical work area, with surgery (n=7; 35%) and resuscitation (n=6; 30%) being the most common (Table 4). OTAS, which was originally developed for surgery, was adapted for use in rounds (O'Leary, Boudreau, Creden, Slade, & Williams, 2012) and handoffs (Nagpal et al., 2011; Symons et al., 2012). Four marker systems were developed to assess behaviors of a single team member: anesthesiologists (Fletcher et al., 2003), emergency medicine physicians (Flowerdew et al., 2012), scrub practitioners (Mitchell et al., 2012a), and surgeons (Yule et al., 2009). Not enough information was available to determine the intended context of measurement in two articles (Capella et al., 2010; Sudikoff, Overly, & Shapiro, 2009).

Scoring Method

Fourteen marker systems reported to use Likert-scales for assessment and 12 of these scales included descriptive anchors at the maximum and minimum values as an assessment aid (Table 4). The number of scale points ranged from three to nine. For example, OTAS ratings cover five behaviors, three subteams (surgical, anaesthetic, nursing), and three operative phases of surgery (pre-operative, intra-operative, post-operative; OTAS, 2011). This results in 45 behavioral ratings for a single surgery. Raters assess performance using a seven point Likert-scale ranging from zero (*Problematic behavior; team function severely hindered*) to six (*Exemplary behavior; very highly effective in enhancing team function*). The Oxford Non-Technical Skills (NOTECHS) scale relies on a summative scoring of behaviors over the entire observation (Mishra, Catchpole, & McCulloch, 2009) Raters assess performance using a four point Likert-scale ranging from one (*Below standard; behavior directly compromises patient* *safety and effective teamwork*) to four (*Excellent; behavior enhances patient safety and teamwork, a model for all other teams*).

Table 4. Overview of target of measurement and scoring method.

Marker System	Scoring Method	Type of Assessment	Number of Scale Points	Temporal Organization	Developed for specific task?	Developed for single team member?	References
Adapted Mayo High Performance Teamwork Scale	Checklist	Presence/ absence of team behavior/ competency	XX	No temporal structure	Y. Resuscitation	N	(Hamilton et al., 2009)
Team Functioning Assessment Tool	Likert scale	Quality of behavior/ competency	7 (N/A option)	Summative assessment of behaviors and categories of behaviors over an entire observed session; Global teamwork score	N	N	(Sutton, Liao, Jimmieson, & Restubog, 2011)
Teamwork Behavioral Rater	Likert-scale; Behavioral rating anchors	Quality of behavior/ competency	7	Summative assessment of behaviors and categories of behaviors over an entire observed session; Global teamwork score	N	N	(Frengley et al., 2011)
Just-in-time pediatric airway provider performance scale (JIT-PAPPS)	Checklist; Weighted	Completion or partial completion of observed event	XX	Sequence of events is temporally based	Y. Airway management	N	(Nishisaki et al., 2011)
Andersen et al. (2010)	Checklist	Presence/ absence of observed event	XX	Events/behaviors are not temporally structured	Y. Cardiac arrest management	N	(Andersen, Jensen, Lippert, Ostergaard, & Klausen, 2010)
Multi-disciplinary Team (MDT) Performance Assessment Tool	Likert-scale; Behavioral rating anchors	Quality of behavior/ competency	5	Summative assessment of behaviors over an entire observed session	Y. Multidisciplinary cancer team meeting	N	(Lamb, Sevdalis, Mostafid, Vincent, & Green, 2011; Lamb, Wong, Vincent, Green, & Sevdalis, 2011)

Marker System	Scoring Method	Type of Assessment	Number of Scale Points	Temporal Organization	Developed for specific task?	Developed for single team member?	References
University of Texas Behavioral Markers for Neonatal Resuscitation (UTBMNR)	Likert scale; Rating anchors	Observability and frequency of ratings	0-5 (observability); 1-4 (frequency)	Summative assessment of observability/frequency of behaviors over an entire session; global scores for teamwork and leadership	Y. Neonatal resuscitation	N	(Thomas, Sexton, & Helmreich, 2004)
Team Emergency Assessment Measure (TEAM)	Likert scale	Quality of behavior/ competency	5 (behaviors); 10 (global)	Summative assessment of behaviors over an entire observed session; 1 global assessment of teamwork	Y. Resuscitation	N	(Cooper et al., 2012; Mullan, Wuestner, Kerr, Christopher, & Patel, 2012)
Observational Skill-based Clinical Assessment tool for Resuscitation (OSCAR)	Likert-scale; Behavioral rating anchors	Quality of behavior/ competency	7	Summative assessment of behaviors and categories of behaviors over an entire observed session (individual and subteam-level)	Y. Resuscitation	N	(Walker et al., 2011)
Simulation Team Assessment Tool	Likert-scale; Behavioral rating anchors	Completion and timeliness of behavior/ competency	3 (N/A option)	Summative assessment of behaviors and categories of behaviors over an entire observed session	Y. Resuscitation	N	(Reid et al., 2012)
Sevdalis et al. (2012)	Frequency count	Quantity of communication frequency	XX	Aggregate assessment of communication types over the course of an observed session	Y. Surgery	N	(Sevdalis et al., 2012)
Objective Teamwork Assessment System (OTAS)	Likert-scale; Behavioral rating anchors	Quality of behavior/ competency	7	Surgical stages: pre-operative, intraoperative, post-operative	Y. Surgery	N	(Hull, Arora, Kassab, Kneebone, & Sevdalis, 2011a; Hull, Arora, Kassab, Kneebone, &

Marker System	Scoring Method	Type of Assessment	Number of Scale Points	Temporal Organization	Developed for specific task?	Developed for single team member?	References
							Sevdalis, 2011b; Mishra et al., 2009; O'Leary et al., 2012; Russ et al., 2012; Symons et al., 2012; Undre, Sevdalis, Healey, Darzi, & Vincent, 2007; Wetzel et al., 2010; Wetzel et al., 2011)
Oxford Non-Technical Skills (NOTECHS)	Likert-scale; Behavioral rating anchors	Quality of behavior/ competency	4	Summative assessment of behaviors and categories of behaviors over an entire observed session	Y. Surgery	N	(Mishra, Catchpole, Dale, & McCulloch, 2008; Mishra et al., 2009)
LOSA Checklist (Adapted)	Likert-scale; Behavioral rating anchors	Quality of behavior/ competency	5	Summative assessment of behaviors, categories of behaviors, and global nontechnical skills over an entire observed session	Y. Surgery	N	(Moorthy, Munz, Adams, Pandey, & Darzi, 2005)
Behaviorally Anchored Rating Scale	Unable to determine	Unable to determine	Unable to determine	Unable to determine	Unable to determine	Unable to determine	(Sudikoff et al., 2009)
Trauma team Performance Observation Tool (TPOT)	Unable to determine	Unable to determine	Unable to determine	Unable to determine	Y. Resuscitation	Unable to determine	(Capella et al., 2010)

Marker System	Scoring Method	Type of Assessment	Number of Scale Points	Temporal Organization	Developed for specific task?	Developed for single team member?	References
Anaesthesia Non-Technical Skills (ANTS)	Likert-scale; Behavioral rating anchors	Quality of behavior/ competency	4 (N/A option)	Summative assessment of behaviors and categories of behaviors over an entire observed session	Y. Surgery	Y. Anesthesiologist	(Fletcher et al., 2003; Graham, Hocking, & Giles, 2010; Westli, Johnsen, Eid, Rasten, & Brattebo, 2010)
Flowerdew et al. (2012)	Likert-scale; Behavioral rating anchors	Quality of behavior/ competency	9	Summative assessment of behaviors and categories of behaviors over an entire observed session	N	Y. Emergency Medicine Physician	(Flowerdew et al., 2012; Flowerdew et al., 2012)
Scrub Practitioners' List of Intraoperative Non-Technical Skills (SPLINTS)	Likert-scale; Behavioral rating anchors	Quality of behavior/ competency	4 (N/A option)	Summative assessment of behaviors and categories of behaviors over an entire observed session	Y. Surgery	Y. Scrub Practitioner	(Mitchell et al., 2012a; Mitchell et al., 2012b)
Non-Technical Skills for Surgeons (NOTSS)	Likert-scale; Behavioral rating anchors	Quality of behavior/ competency	4 (N/A option)	Summative assessment of behaviors and categories of behaviors over an entire observed session	Y. Surgery	Y. Surgeon	(Crossley, Marriott, Purdie, & Beard, 2011; Yule, Flin, Paterson-Brown, Maran, & Rowley, 2006; Yule et al., 2009)

In addition to Likert-scales, three marker systems relied on checklists and one marker system used a frequency count. Andersen and colleagues (2010) developed a checklist with 22 behavioral markers for the formative assessment of resuscitation teams. Raters identify the occurrence or absence of target behaviors during assessment, but there is no chronological sequence for when raters can expect behaviors to occur. Conversely, the JIT-PAPPS checklist uses a temporal structure to assess whether certain actions during airway management simulations were accomplished, partially accomplished, or not done all. These actions were linked to competencies such as situational awareness (SA), decision-making, task management, and teamwork. Further, certain tasks were weighted to connote heightened importance of a particular skill.

Evidence of Reliability and Validity

The inferences drawn from measurement must be considered in relation to the established psychometric properties of a measurement tool. The reliability of a measure concerns its consistency over repeated measurements. Validity addresses its accuracy and utility in relation to the performance context and the quality of inferences that can be made from a specific process of data collection (Pedhazur & Schmelkin, 1991). Establishing the reliability of a measure is necessary, but not sufficient for ensuring its validity (Cook & Beckman, 2006).

Types of reliability and validity reported in the 37 articles are described in Table 5 and related evidence is summarized in Table 6. Reliability evidence was reported for 15 marker systems and evidence of validity was reported for 14 marker systems. Content validity was the most common type of validity evidence presented (n=11; 55%), followed by observability (n=5; 25%), concurrent/convergent validity (n=4; 20%) and sensitivity (n=4; 15%). Crossley and colleagues (2011) discussed a comprehensive process to evaluate the reliability and validity of

NOTSS. This evaluation involved 715 assessments of 404 surgical cases by four types of raters who had received minimal training (56 anaesthetists, 39 scrub nurses, 2 surgical care practitioners, and 3 independent raters). The authors applied generalizability (G) theory to demonstrate good reliability for the marker system. NOTSS scores were also subjected to exploratory factor analysis to establish the internal structure of the marker system. The solution demonstrated a pattern of results mostly consistent with the hierarchical structure specified by the instrument with the exception of the behavior 'setting and maintaining standards' which loaded on both Leadership and Situation Awareness. The relationship of NOTSS categories to external variables was also examined to confirm NOTSS measured attributes related to nontechnical and training-related skills.

Mitchell et al. (2012b) examined the reliability and validity of the Scrub Practitioners' List of Intraoperative Non-Technical Skills (SPLINTS) system. Thirty-four scrub practitioners attended a one day training session in the use of SPLINTS and then rated nontechnical performance in seven standardized video simulations. Evaluation criteria focused on reliability (within-group agreement and internal consistency), validity (accuracy, completeness, observability), and usability (acceptability and usability). Within-group agreement was good for each skill category, but one-third of skill elements did not reach acceptable thresholds ($rwg > 0.7$). Within-group agreement also varied by scenario. Estimates of internal consistency corroborated the hierarchal structure of the measurement system and participants indicated that SPLINTS addressed important nontechnical skills and that behaviors were easy to observe.

Table 5. Types of reliability and validity evidence reported.

Types of Reliability Reported	Definition
Interrater reliability	The relationship of rating scores between two or more assessors for the same attribute over multiple rating periods (Trochim, October 20, 2006).
Interrater agreement	Consistency in which raters assess an attribute with higher (or lower) scores across rating periods.
Internal consistency	Estimates how well items and/or subparts of a measurement instrument actually measure the attribute they are purported to measure (Pedhazur & Schmelkin, 1991).
Test-Retest Reliability	The consistency of a measurement instrument across multiple rating periods (Trochim, October 20, 2006).
Generalizability Theory	Generalizability (G) and Decision (D) studies are conducted to identify the magnitude and sources of measurement error and estimate the dependability of a measurement instrument with alternative research designs (Brennan, 2001).
Types of Validity Reported	**Definition**
Content Validity	The extent to which items of a measurement instrument are important and relevant to a performance context (Lynn, 1986).
Convergent/ Concurrent Validity	The degree to which constructs of a measurement instrument are correlated with similar approaches purported to measure the same construct (Campbell & Fiske, 1959).
Observability	The extent to which behaviors being targeted for measurement can actually be observed by raters (Fletcher et al., 2003) .
Relationship to external variables/ Sensitivity	Different conditions should elicit unique responses and the marker system should be able to make this discrimination.
Completeness	The scope of the measurement instrument is comprehensive and captures all relevant behaviors (Mitchell et al., 2012b).
Accuracy	The ability of raters to make accurate assessments (Fletcher et al., 2003). Accuracy differs from interrater reliability because raters can be consistent, but not accurate.
Internal Structure	Exploratory factor analysis is applied to compare how well a set of indicators of a construct match the organizational structure explicated by the measurement instrument (Tabachnick & Fidell, 2007).
Unidimensional validity	The item to total score correlation for a measurement instrument (Cooper et al., 2012).

Table 6. Evidence of reliability and validity.

Marker System	Evidence of reliability	Evidence of validity
Andersen et al. (2010)	1. Interrater reliability: ICCs =.9 (95% CI: 0.79–0.97) (Andersen et al., 2010). 2. Single item agreement: rate of instructor agreement on single items of checklist ranged from 0.58 to 0.82. (Andersen et al., 2010) 3. Single item agreement: Kappa for single items ranged from 0.03 to 0.82. (Andersen et al., 2010)	1. Content validity: interviews to determine initial needs and presentation to a group of ALS instructors/providers (Andersen et al., 2010) . 2. Concurrent validity: compared instructor scores to reference values (ICC = .93; 95% CI: 0.71–0.98) (Andersen et al., 2010).
ANTS	1. Interrater reliability: ICCs = .72 ($p<.05$) (for trauma training) (Westli et al., 2010). 2. Interrater reliability: Rater agreement between expert raters and trainee raters ranged from 0.11 to. 062.(Graham et al., 2010) 3. Internal consistency: Cronbach's alpha for each dimension ranged from 0.77 to 0.87;(Graham et al., 2010) (Chronbach alpha ranged from 0.79 to 0.86.(Fletcher et al., 2003) 4. Interrater agreement: Ranged from 0.55 to 0.67 at the element level and 0.56 to 0.65 at the categorical level (Fletcher et al., 2003).	1. Completeness: (survey of participants) 100% of participants indicated that ANTS addressed key non-tech skills and 84% did not feel any ANTS elements were missing from videos (Fletcher et al., 2003). 2. Observability (Survey of participants): Elements were observable greater than 80% of the time and categories were observable more than 95% of the time (Fletcher et al., 2003). 3. Accuracy/sensitivity: Average deviation from referent ratings (Fletcher et al., 2003). 4. Content Validity: Cognitive task analysis reported from previous development (Fletcher et al., 2003).
Behaviorally Anchored Rating Scale	Not Specified.	Not Specified.
Flowerdew et al. (2012)	1. Interrater reliability: ICCs for 3 pairs of observers was low (0.575, 0.532 and 0.419) (Flowerdew et al., 2012). 2. Test-retest reliability: Spearmans Rho was 0.26 when examining individual skills but mean scores were 0.70 (Flowerdew et al., 2012).	1. Content validity: Assessed using content validity index 0.75 (participants rating an item as very important/essential divided by the total number of participants) (Flowerdew et al., 2012). 2. Observability: Frequency of skills observed divided by number of assessments (Flowerdew et al., 2012).
Just-in-time pediatric airway provider	1. Interrater reliability: Correlation coefficient between expert rater and RAs for the overall (.73, $p = .001$; .88, $< .001$) and behavioral domain (0.63, $p=.009$; .84 $p <.001$)	Not Specified.

Marker System	Evidence of reliability	Evidence of validity
performance scale (JIT-PAPPS)	(Nishisaki et al., 2011).	
LOSA Checklist (Adapted)	1. Interrater reliability: Alpha = 0.84 (Moorthy et al., 2005).	Not Specified.
Mayo High Performance Teamwork Scale (Adapted)	Not Specified.	1. Sensitivity to skill: Capability of rating system to discriminate between effective/ineffective performance (p<.001) (Hamilton et al., 2009). 2. Sensitivity to scenario: raters identified the presence of certain team attributes differently based on the scenario (53% to 94%, p<.001) (Hamilton et al., 2009).
MDT Performance Assessment Tool	1. Interrater reliability: ICCs ranged from 0.31 to 0.87 (Lamb et al., 2011).	1. Concurrent validity: Median correlation was rho = 0.74 between observer's ratings and self-reported scores (Lamb et al., 2011). 2. Content validity: Review of team performance and assessment tools and adaption of existing assessment tools (Lamb et al., 2011). 3. Face validity: An oncologist and a CNS were consulted (Lamb et al., 2011).
NOTECHS	1. Interrater reliability: Rwg = 0.99;(Mishra et al., 2009) Chronbach's alpha = 0.880 (Mishra et al., 2008). 2. Test-retest reliability: non-significant differences observed between the 3 intervention periods (p=0.281) and post intervention periods (p=0.368) (Mishra et al., 2009).	1. Sensitivity to performance differences: Capability of rating system to explain performance differences: rho=-0.413, n=65, p=.001) (Mishra et al., 2009). 2. Convergent validity: Comparison between OTAS and NOTECHS ratings (r=0.886, n=5, p=0.046) (Mishra et al., 2009).
NOTSS	1. G and D studies demonstrated good reliability (Crossley et al., 2011). 2. Interrater agreement: the mode rating of non-tech behaviors was the same as the expert group half of the time (12/24) (Yule et al., 2009).	1. Internal Structure: Exploratory factor analysis (Crossley et al., 2011). 2. Relationship to external variables: Intercorrelations with other measures (Crossley et al., 2011). 3. Content validity: Cognitive task analysis (Yule et al., 2006).

Marker System	Evidence of reliability	Evidence of validity
OSCAR	1. Internal consistency: Cronbach's alpha for all behaviors were greater than 0.70 (Walker et al., 2011). 2. Interrater reliability: ICCs were high for all behaviors for each subgroup of clinicians (Walker et al., 2011).	1. Face validity: 10 experts rated validity, 39 behaviors were rated as critically important (Walker et al., 2011). 2. Content validity: Same as face validity (Walker et al., 2011).
OTAS	1. Interrater reliability: Rho ranged from 0.53 to 0.68 for each category (rounds) (O'Leary et al., 2012); ICC = 0.61 (P<0.001) (handoff) (Symons et al., 2012); Spearman correlation = .829 (p<.001) (handoff) (Nagpal et al., 2011); Significant correlation between raters (r=.71, p<.01) (Hull et al., 2011a); Rater scores were highly correlated with each other for observed behaviors (r >0.50), with the exception of communication (r=0.35), (Undre et al., 2007). 2. Interrater agreement: Mean ICCs between expert/novice raters. Significant improvement in interrater reliability over time F(2,27) = 4.12-22.95, p's < 0.05 (Russ et al., 2012).	1. Content validity: High user involvement in Delphi sampling (for handoff) (Nagpal et al., 2011). 2. Content validity: Exemplars were refined by a panel of experts (Hull et al., 2011b). 3. Observability: (Interobserver agreement of presence/absence of 130 behaviors) Cohen's K was greater than or equal to 0.41 for 84% of behaviors (Hull et al., 2011b). 4. Concurrent validity: Correlation with another teamwork measure (for handoff) (Symons et al., 2012).
Sevdalis et al. (2012)	Interobserver Agreement: ICC=0.92–0.98 (P = 0.001) (Sevdalis et al., 2012).	Not Specified.
SPLINTS	1. Interrater agreement: Within-group agreement (rwg > 0.70) (Mitchell et al., 2012b). 2. Internal consistency: mean absolute difference < 0.2 of a scale point between category and elements (Mitchell et al., 2012b).	1. Content validity: Focus groups (Mitchell et al., 2012a). 2. Completeness: Degree to which participants feel SPLINTS addressed requisite nontechnical skills (Mitchell et al., 2012b). 3. Observability: Participants determined it was (a) very easy, (b) easy, or (c) average amount of ease to link behaviors with elements (Mitchell et al., 2012b). 4. Accuracy/Sensitivity: Average deviation from referent ratings (Mitchell et al., 2012b).

Marker System	Evidence of reliability	Evidence of validity
STAT	1. Interrater reliability: ICC=.81(Reid et al., 2012).	1. Content validity: Review of tool by 7 experts (Reid et al., 2012). 2. Sensitivity to skill: Capability of STAT to distinguish expert/ novice performance: Experts performed significantly better (mean = .84) than novices (mean = .66, p =0.02) (Reid et al., 2012).
TEAM	1. Internal consistency: a = .923 (Cooper et al., 2012).	1. Unidimensional validity: Item to total scale correlations ranged from 0.583 to 0.909 (Cooper et al., 2012).
Teamwork Behavioral Rater	Not Specified.	Not Specified.
TFAT	1. Interrater reliability: Kendall's coefficients of concordance W were all significant for each subscale as well as overall score (Sutton et al., 2011).	1. Content validity: Card sorting exercise assessing whether experts could determine and sort the 15 behavioral elements into the five core categories (Sutton et al., 2011).
TPOT	Not Specified.	Not Specified.
University of Texas Behavioral Markers for Neonatal Resuscitation	Not Specified.	1. Content validity: Focus groups (Thomas et al., 2004). 2. Observability: Review of video recordings of resuscitations of infants (Thomas et al., 2004).

The majority marker systems reported evidence of content validity and several approaches for establishing content validity were applied. Flowerdew et al. (2012) surveyed hospital staff to rate the degree to which certain nontechnical skills were applicable to an emergency department and created an index of content validity (i.e., the proportion of total respondents rating a behavior as important). Nagpal et al. (2011) used Delphi sampling—a consensus building technique— to establish the content of the Postoperative Handover Assessment Tool (PoHAT), which leveraged OTAS dimensions to score teamwork.

In sum, interrater reliability was the most widely cited index of measurement consistency. A variety of approaches have been used to assess validity, including: convergence with other rating systems (Lamb et al., 2011), sensitivity to scenario attributes (Hamilton et al., 2009), content of the measurement instrument (Flowerdew et al., 2012; Walker et al., 2011), and the completeness and observability of the marker system (Fletcher et al., 2003).

Required Skills and Training

Calibrating rater scores is necessary to ensure research results are reliable, which is generally achieved through rater training. Information detailing the length of rater training was reported in 27% of articles (n=10). The time spent training raters ranged from just over two hours (Yule et al., 2009) to over two days (Russ et al., 2012). Four separate marker systems reported the length of rater training to last at least one day (Frengley et al., 2011; Graham et al., 2010; Russ et al., 2012; Sutton et al., 2011) and other articles simply stated that rater training was 'minimal' (Crossley et al., 2011) or involved a certain amount of practice observations of unspecified structure or duration (O'Leary et al., 2012).

Ratings made between novice raters and expert referents demonstrated good reliability in as little as four to six hours of training (Fletcher et al., 2003; Mitchell et al., 2012b) while other examples were much more time-intensive, lasting over two days (Russ et al., 2012). Russ et al. (2012) reported how the reliability between expert and novice ratings using OTAS improved at each stage of rater training, with the learning curve being contingent upon the construct being measured. Rater training involved approximately two hours of declarative information presentation followed by one hour of video-based practice. Next, raters observed 10 surgical cases and received immediate feedback on their assessments during post-observation debriefings (approximately 18 total hours). High rater calibration for coordination was established immediately so improvements were not significant due to a ceiling effect. Considerable improvements were demonstrated for communication, cooperation, and leadership over the first seven observations while steady improvements in rater calibration for monitoring/SA were demonstrated over the entire observation period. Further, there was no significant difference between novice raters with different professional backgrounds (i.e., surgery and psychology).

The impact of rater training on rater performance was mixed, however. Following a two and a half hour NOTSS training course, the mode rating of nontechnical behaviors made by novices was the same as experts only half of the time (Yule et al., 2009). Additionally, novices tended to underrate nontechnical performance compared to experts (Yule et al., 2009). Graham et al. (2010) found considerable differences between expert and novice ratings following a one day ANTS training session, with a major source of disagreement being the misclassification on nontechnical skills; raters were identifying behaviors, but scoring them as different elements of teamwork. Finally, Lamb et al. (2011) reported a significant difference in ratings made between disciplines (i.e., surgeon and psychologist), though there was a significant improvement of intraclass correlations (ICCs) as more cases were observed.

Applications of Behavioral Marker Systems in Healthcare Research

Fifteen articles employed marker systems to test the relationship between constructs (n=4), study the effects of an intervention (n=5), or describe teamwork in relation to task activities (n=7). Westli et al. (2010) investigated team skills during trauma simulations. Team membership included a surgeon as the team lead, an anesthesiologist, an anaesthetic nurse, an emergency medical nurse, and a radiographer. Positive relationships between performance and competencies such as information exchange, coordination, communication, and SA were reported. Surprisingly, higher performing teams demonstrated less supporting behavior. Other studies reported differences in teamwork scores based on professional background (Hull et al., 2011a; Mishra et al., 2008) and years of experience (Wetzel et al., 2010).

Behavioral marker systems have also been employed to establish the effectiveness of training interventions (Capella et al., 2010; Nishisaki et al., 2011). Frengley et al. (2011) evaluated the relative effectiveness of simulation-based training and case-based learning on the management of airway and cardiac crises with the Teamwork Behavior Rater. The authors reported teamwork skills significantly improved for both intervention strategies. Feedback/debriefing on teamwork skills during training was described in four articles, but none described the process of how feedback was delivered or whether it was structured.

With respect to task activities, Symons et al. (2012) adapted OTAS to study teamwork skills in a handoff. Despite establishing adequate interrater reliability and concurrent validity with another teamwork scale, the authors did not observe significant correlations between teamwork skills and the completion of handoff content, handoff length, interruptions during handoffs, or attendance at handoffs. Sevdalis et al. (2012) observed that communication events were most likely initiated by surgeons (80%) and were received by either surgeons (46-56%) or nurses (38-40%). Additionally, laparoscopic surgeries tended to involve more communication events that were equipment related and that were directive compared to open surgeries. Another study found surgeons' SA was negatively correlated with technical errors (Mishra et al., 2008).

Forty percent of applied research articles did not report evidence of rater reliability or training.

Implications

This chapter answered four research questions surrounding the use of behavioral marker systems in healthcare. First, this chapter intended to identify the attributes of behavioral marker systems. A surprisingly large number of unique skills and competencies were found to be targeted for measurement. It is likely that marker systems cover similar content, but inconsistent terminology and differing levels of granularity used to describe constructs complicates the comparison of behavioral marker content across systems. This finding is consistent with a previous review of medical teamwork (Baker et al., 2005; see also Chapter 2) as well as reviews of marker systems in other domains (Flin & Martin, 2001). Additionally, the majority of marker systems were developed for a specific work domain or task, yet none were specific to ICU teams.

The temporal structure or resolution of a measurement system is a key attribute with implications for ease of training and data use. Most marker systems used a low resolution time scale where behavioral assessments were made once over the entire rating period (e.g., a team received one score for a dimension for the entire observation period). Low temporal resolution ratings may illuminate *what* teamwork deficiencies exist, but not necessarily *why* they occurred (Salas, Rosen, Burke, Nicholson, & Howse, 2007). Conversely, systems with higher levels of temporal resolution identify phases of performance or multiple time blocks within an observational period. For example, OTAS rates teamwork dimensions across three phases of surgery and JIT-PAPPS used an event based approach to measurement (EBAT) (Nishisaki et al., 2011). EBAT tools rate teamwork competencies and skills relative to stimulus events (Dwyer, Fowlkes, Oser, Salas, & Lane, 1997; J. Fowlkes, Lane, Salas, Franz, & Oser, 1994; J. E. Fowlkes & Burke, 2005; Rosen et al., 2008). This approach is most useful for training or for assessments of tasks that are highly structured, where scripted scenario events provide opportunities for trainees to exhibit teamwork skills (Rosen et al., 2010). These systems are viable for providing explicit feedback on processes that explain *why* deficiencies in teamwork may exist. Further, this approach may reduce the cognitive load placed on raters by explicating what is supposed to be assessed and when; raters detect the presence or absence of events following an observation checklist that is temporally constructed which can enhance objectivity (Flin & Martin, 2001). A key shortcoming to EBAT is that generalizability is limited to the context and task being assessed. For instance, stimulus events indicative of teamwork skills for a resuscitation task would be fundamentally different for a handoff.

A second objective of this chapter was to examine evidence of reliability and validity. The most widely cited index of reliability was the calibration of scores among raters, yet interrater reliability only estimates one source of measurement error: the rater. In reality, error variance and systematic bias in ratings can come from of other sources such as the time of observation, participants being observed, and the context of observation. Unlike traditional approaches to reliability testing, which estimate a single source of measurement error, G studies seek to categorize and explain the magnitude of error variance for multiple sources (Brennan, 2001; DeShon, 2002). This information is then used to determine whether alternative designs would minimize measurement error in future applications (Shavelson & Webb, 1991). Although G theory is a valuable approach for reliability testing, it was only carried out in one study

(Crossley et al., 2011). G theory provides a future opportunity for researchers to unequivocally define and account for sources of measurement error. Therefore, the present study will apply G theory to provide evidence of reliability and validity.

Marker systems seek to quantify abstract variables (e.g., cooperation, SA) with observed behaviors, which makes the quality of inferences that can be drawn from these tools of paramount importance. The most prevalent approach to establish validity was to authenticate the veracity of measurement content (i.e., content validity). Evidence of validity should come from a variety of sources, such as the content of the measurement tool, whether or not team competencies are actually being observed, the tools' internal structure, and convergent and discriminate relationships with other constructs to name only a few (Osterlind, 2010). Clearly, extensive evidence is needed to establish the validity of a marker system, but multiple sources of evidence were only reported for 12 marker systems. While exceptions exist (Table 6), the marker system literature requires further validation research. This finding is consistent with previous reviews on performance measurement in healthcare (Jeffcott & Mackenzie, 2008). Therefore, an objective of the current study is to present multiple sources of reliability and validity evidence.

The third objective of this chapter was to understand what expertise is required for raters to sufficiently judge performance. Accurate judgments of behaviors directly influence the validity of inferences drawn from measurement and all raters are susceptible to biases, no matter their professional background. This makes rater training necessary to immerse raters in the content of the marker system, its appropriate use for observation, and to curtail the possibility of rater biases manifesting during assessments (Flin & Martin, 2001). Best practices for behavioral assessment call for recurrent rater training and reliability testing to ensure rater scores are calibrated and accurate over time (Rosen et al., 2010). The impact of rater training on rater proficiency was mixed (Mitchell et al., 2012b; Nishisaki et al., 2011; Yule et al., 2009), but evidence suggests improvements can be made over time (Russ et al., 2012).

The final aim of this chapter was to examine the application of behavioral marker systems in healthcare. Focal shortcomings identified in this review were deficiencies in rater training and reliability reporting. Just because a measure has demonstrated evidence of reliability and validity in previous research does not mean it will inevitably be successful in a new context (Healey, Undre, & Vincent, 2004). Raters must still be trained in the use of the measurement system and reliability testing should be reported to ensure the veracity of conclusions.

At this point, it is worth acknowledging a particularly relevant article that was not returned in the search results for review and synthesis (see Figure 3, Stage 1). Weller and colleagues (2011) developed a marker system to measure ICU teamwork and applied G theory in their reliability assessment. The focus of measurement, however, was restricted to four emergency scenarios and the G study was conducted in a simulated research setting. The content of the marker system was also adapted from the Mayo High Performance Teamwork Scale, with the authors revising and adding items as needed. Like other marker systems reviewed, the authors combined unique teamwork competencies as a single rating dimension: 'Leadership' and 'Team Coordination.' The present study seeks to develop and evaluate a much more comprehensive marker system, with content methodically developed through literature syntheses (Chapters 2 and 3), subject matter interviews (Chapter 4), and tested in a naturalistic setting (Chapters 5 and 6).

Conclusions

Key features of behavioral maker systems in healthcare were reviewed to guide the development of the ICU system for this study. In general, behavioral marker systems are designed for specific work domains or tasks. Marker systems vary in their content and structure, even for the same task. Although several approaches were applied to establish reliability and validity, the marker system literature as a whole requires more robust reliability and validity evidence. Research considerations for the current study include: (1) establishing evidence of reliability and validity from multiple forms (a single index is not sufficient), (2) applying G theory in addition to traditional reliability metrics, (3) reporting how raters were immersed in the content of the measurement tool, and (4) selecting a scoring format appropriate for the tasks being rated (e.g., high vs. low temporal resolution).

CHAPTER 4: DEVELOPING AN ICU BEHAVIORAL MARKER SYSTEM

Chapters 2 and 3 illuminated a number of critical issues that support the need to develop a theoretically-based, psychometrically-driven behavioral marker system to assess ICU team performance. As reported in Chapter 1, "medical teamwork and team training research are not formally linked to medical team performance theory" (Baker et al. 2005, p.50). Nearly 10 years since this conclusion, findings from Chapter 2 demonstrated that theoretical shortcomings are still prevalent in the ICU teamwork literature, though notable exceptions exist (e.g., Reader et al., 2009).

Findings from both systematic literature reviews revealed that the terms used to describe teamwork dimensions can vary in research studies, offering little confidence the reported findings are manifestations of constructs under investigation. Even more concerning is the mislabeling of constructs as noted in the behavioral marker review. For instance, a marker system that includes a rating dimension of 'Teamwork and Cooperation' inherently confuses the inferences that can be drawn from that dimension of performance measurement: cooperation is an aspect of teamwork, which encompasses other knowledge, skill, and attitude components (see Chapter 1). There is also a dearth of adequate reliability and validity evidence across the marker system literature as a whole and those marker systems with sufficient reliability and validity evidence were developed for either a single team member or team task. Further, because contextual and task related factors are likely to dictate which teamwork competencies are most important, existing teamwork assessment strategies are not guaranteed to generalize across domains of work or research (Healey et al., 2004). These findings suggest an existing marker system cannot be readily adapted for the ICU.

Mapping Teamwork Theory with Context-Specific Exemplars

When conceptualizing a behavioral marker system for ICU team performance, good practice dictates the importance of sound theoretical underpinnings to guide measurement (Baker & Salas, 1992). At the same time, context-specific direction is needed to provide valid accounts for how indicators of teamwork competencies uniquely manifest in this performance context. Therefore, this chapter describes how teamwork theory was mapped with context-specific exemplars to identify a framework of the focal factors that drive ICU team performance. This intersection of top-down (i.e., theoretical) and bottom-up (i.e., context-driven) guidance serves as an indicator of the validity of the behavioral marker system's content while simultaneously ensuring operationalizations of teamwork dimensions (i.e., specific behavioral markers) are formally linked to teamwork theory. Table 7 provides an overview of the marker system's development.

Table 7. Overview of ICU behavioral marker development.

	Key Points
Purpose	• This chapter describes how teamwork theory was mapped with context-specific exemplars to identify a framework of the focal factors that drive ICU team performance.
Methods	• Critical incident interviews with ICU clinicians (11 nurses and 9 physicians) were conducted to identify conditions where teamwork was particularly effective (or not) and the underlying factors driving performance (or poor performance).
Results	• ICU teams, like all teams, cycle through transition and action phases of teamwork. During each of these phases, some teamwork competencies are more salient than others. • Four core dimensions were identified that provide the theoretical foundation for the ICU marker system: communication (global), leadership (global), backup and supportive behavior (action), and team decision-making (transition), • Behavioral markers extracted from critical incident interviews were mapped to these teamwork dimensions to determine fit.

Critical Incident Interviews

Data for this component of the study were collected by the author as a Senior Research Program Coordinator II for the Johns Hopkins University (JHU), School of Medicine as part of a much larger research initiative (a subset of interviews were conducted by the author and a human factors psychologist). The Institutional Review Board (IRB) at JHU granted approval for these interviews. The findings presented in this section report relevant data captured from the critical incident interviews only as they relate to the current study.

Twenty clinical team members from two ICUs (11 nurses and 9 physicians) were recruited to participate. Participants were informed of the purpose of the study, possible risks associated with data collection, and asked to orally consent to participate before and after each interview. The interviews lasted approximately 30 minutes and participants were asked a series of open-ended questions about the competencies they felt drive team performance and conditions when teamwork matters most (see Appendix C for protocol). Additionally, participants were asked to walk through specific events in which breakdowns in team processes occurred, scenarios where team performance was particularly effective, and the circumstances surrounding those events using the critical incident technique (Flanagan, 1954). Interview findings were transcribed by the author. Although interviews were not documented verbatim, an exhaustive account from each interview was captured for content analysis.

Team Tasks

A summary of the major types of ICU team tasks described by participants is presented in Table 8. The science of teamwork explicates that teams cycle between action and transition phases of team task accomplishment (LePine et al., 2008; Marks et al., 2001) and ICU teams are no different (see Figure 4). In the ICU, transition cycles involve conveying care priorities, developing goals, and formulating strategies for achieving those goals. Transition cycles occur during planned clinical activities such as rounds and handoffs as well as emergently throughout the day depending on changes in patient conditions, or updates to previously specified plans (e.g., scheduling tests and procedures, acquiring resources for transfer or admission). Action cycles refer to those activities directly involved in providing health services to the patient, such as administering medications, conducting procedures, or responding to a code event. Figure 5 illustrates a representative sample of team tasks listed in Table 8 in relation to action and transition cycles of team performance.

As described by Marks et al. (2001), team processes may be more salient depending on the cycle of team task accomplishment. For instance strategy formulation and planning is particularly relevant during transition cycles while team monitoring and backup are central to action phases. Last, interpersonal processes such as conflict management, motivation and confidence building, and affect management transcend action and transition phases.

Absent from the classification of teamwork processes outlined by Marks and colleagues (2001) are additional competencies that are globally relevant across action and transition phases (i.e., other than interpersonal processes). For example, communication and leadership were highly cited in the systematic literature review of teamwork in the ICU (Chapter 2) as well as the critical incident interviews; they represent important components of teamwork for assessment regardless of the cycle of team task accomplishment. In reality, ICU team performance is contingent on competencies that are either specific or generic in relation to tasks that are performed. As described by Cannon-Bowers et al. (1995) generic competencies are important independent of the task that is performed (e.g., backup behavior). Conversely, task specific competencies are only germane to a particular team task (e.g., cue strategy associations).

Table 8. An overview of ICU team tasks identified in critical incident interviews.

Task	Description
General Report (Nursing)	• Oncoming nursing team members review the patients currently on the unit, discussing the reason for their admission and their acuity level and receive patient assignments during this time (typically 1-2 patients per nurse).
Handoff (Shift Report)	• Nursing team members receive a report from the outgoing nurse on the patient(s) they were assigned. This report includes a systematic head-to-toe assessment of the patient by organ system. Additionally, they should cover why the patient was admitted to the ICU, major clinical activities that have taken place, what the plans are, and what the patient's hopes are. • Physician handoffs cover a larger subset of the patients and may be a formal or informal team activity depending on the unit. Physician team members solicit information from other clinical team members (e.g., fellow-to-fellow, fellow-to-resident, resident-to-resident, fellow-to-nurse, and resident-to-nurse) and through patient databases (e.g., electronic medical record, charts) about key events that have taken place, medication titrations, planned activities, and trends in patient vitals.
Rounds	• Goals of care and methods of task accomplishment for each patient are established. • Rounds typically last 15 to 30 minutes per patient, and involve: o A report from nurses on new or acute issues; an objective case presentation from residents, followed by a subjective interpretation, their recommendation, and plan; and a team discussion of the merits of the proposed plan and possible alternate treatment options/contingencies. • Rounds are managed by either the attending or fellow (depending on the attending's leadership preference). The attending or fellow that is facilitating rounds specifies who is responsible for documenting meds and putting in orders. • The attending and/or fellow also integrate teaching opportunities into the discussion of care plans and fills in any gaps omitted from the case presentation.
Running the List	• Physician team members deliberately review and discuss the planned course of patient care for each patient following rounds. • The fellow confirms that the physician team has a shared understanding of care plans and priorities and delegates tasks for the resident to complete.
Admissions	• Clinical team members prepare for and manage new arrivals to the unit. Physician team members take report (i.e., the handoff) and begin writing orders. Nursing team members help get the patient situated, which involves activities such as documentation, getting the patient hooked-up to monitors, and doing a full assessment (e.g., lungs, mental status).
Routine Care/ Procedures	• Team members execute the plans that were developed during morning rounds (e.g., lab work, line placement, scheduling tests and temporary unit transfers, sedation interruptions). • Team members also conduct routine care activities. For instance: o Patients need to be repositioned every 2 hours if they are unable to do so themselves. o Oral care should be administered every 4 hours for patients that are vented. o Team members need to conduct routine assessments of patient vitals. o Physician team members place orders emergently throughout the day.
Code/ Emergency Event	• Team members rapidly respond to a patient that is coding. A single physician takes charge of leading the code (typically the fellow or attending) and delegates roles/responsibilities for other team members to complete.
Discharge	• Patients that are stable enough to be transferred to the floor (or from the hospital) are discharged from the unit. • Intra-unit activities related to discharge include discussions among the attending, fellow, and charge nurse to verify which patients are expected to be discharged from the unit.

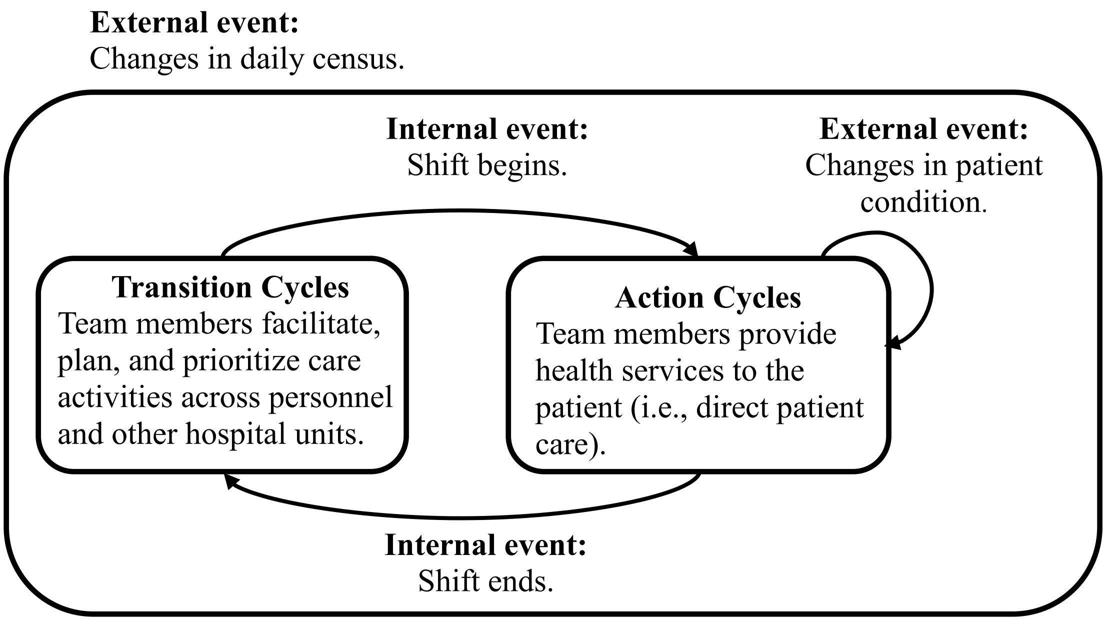

Figure 4. Cycles of ICU team task accomplishment.
(c.f. Marks et al., 2001).

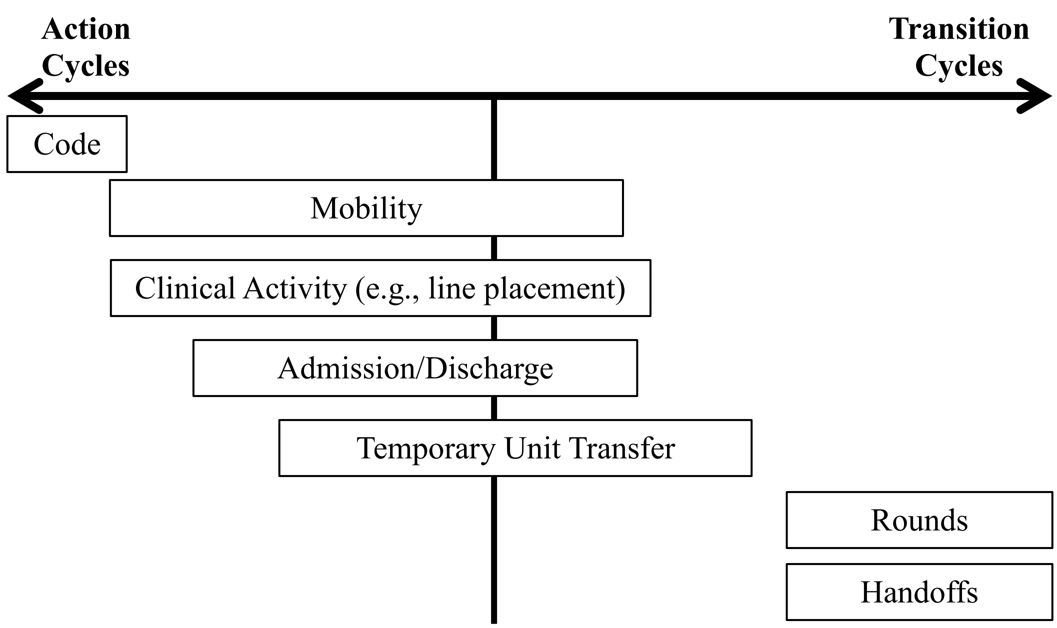

Figure 5. Team tasks in relation to cycles of team task accomplishment.
(c.f. Marks et al., 2001, p. 364).

Team Competencies

A key challenge to selecting competencies to target for measurement involves the delicate balance between the specificity of measurement content and factors related to accessibility and generalizability. The panoply of possible constructs to target for measurement is apparent from the review on ICU teamwork in Chapter 2. It is unrealistic, however, to ask an observer to rate performance on such an array of competencies in a single assessment period; the theoretical framework guiding measurement should be parsimonious yet impactful. Therefore, team competencies targeted for measurement should focus on the most critical aspects of teamwork in relation to the performance environment. It is also worth noting that existing ICU team performance classification structures did not lend well to instrument development for the current study. Specifically, the framework presented by Reader and colleagues (2009) identified four facets of teamwork germane to ICUs: Team Communication, Team Leadership, Team Coordination, and Team Decision-Making. The definitions of these terms, however, proved too broad for classifying behavioral markers extracted from interview findings (see section below) in relation to these competencies. Therefore, the definitions of teamwork constructs identified for the ICU behavioral marker system must be explicit enough to ensure behavioral indicators are classified correctly during measurement. That is, markers should 'load' on a single competency as described in the NOTSS validation effort reported in Chapter 3 (Crossley et al., 2011).

Based on findings from the ICU team-based literature review (Chapter 2), science of teams, and insights from the critical incident interviews, four thematic content areas that capture core teamwork competencies of ICU teams have emerged (see Table 9). As illustrated in Figure 6, *Communication* and *Leadership* are globally relevant to ICU team performance (i.e., these competencies are relevant across action and transition cycles of team task accomplishment).

Conversely, *Backup and Supporting Behaviors* are more likely to occur during action phases of teamwork, where team members must reallocate work in relation to shifting priorities or seek out help in order to complete a task. Given the emphasis on developing plans and strategies to accomplish care goals in the ICU, *Team Decision-Making* was also identified as a focal component of teamwork. In fact, the physician team can spend as much as three to four hours of their day in daily rounds (a key decision-making task). Following rounds, the rest of the day is focused on implementing those care plans (e.g., tasks classified as action phases) and updating other team members on the progress of care goals to ensure patient care objectives are realized. This framework is consistent with existing conceptualizations of teamwork in the ICU (Reader et al., 2009) and medical teamwork in general (e.g., TeamSTEPPS; American Institutes for Research, 2010), yet differs due to its explicit definitions of teamwork constructs and emphasis on teamwork skills that are especially relevant to ICU teams.

The subsequent section will detail how critical incident interviews were conducted to elicit positive and negative examples of teamwork. Analyzing the content of these interviews in relation to the competencies described above will ensure the appropriate fusion of context-specific behaviors with teamwork theory. Figure 7 illustrates a framework of marker system development to this end. Specifically, this example shows how behavioral indicators for *Communication* could be derived for daily rounds (a transition cycle of teamwork). Of interest in this example is emphasis on patient-centered care (Carayon & Friesdorf, 2006). To illustrate, the use of terminology among clinicians when specifying care plans will be different amongst themselves as opposed to a laymen's characterization of care plans to ensure patient and/or family member understanding. Thus, behavioral markers of communication are likely to be different for clinician-to-clinician interactions as opposed to clinician-to-patient interactions.

Table 9. A framework of ICU team performance.

Dimension	Definition	Citation
Communication	Communication refers to the style and structure of how information is conveyed between team members. Communication entails exchanging messages using standardized protocols with appropriate terminology in a manner that is clear, accurate, and succinct. A key feature of communication exchanges is that they are closed-loop; the sender conveys information, the receiver confirms the receipt of information, and the sender clarifies any misunderstandings.	(Cannon-Bowers et al., 1995; Reader et al., 2009; Salas, Rosen, Curke, Nicholson, & Howse, 2007; Smith-Jentsch, Zeisig, Acton, & McPherson, 1998)
Leadership	Team leadership refers to the management of team resources/personnel, establishment of team norms, and provision of opportunities to foster the development of knowledge and skills. Team leaders ensure there is clarity of team member roles/responsibilities and that input from all team members is welcomed.	(Reader, Flin, & Cuthbertson, 2011; Salas, Sims, & Burke, 2005; Zaccaro, Rittman, & Marks, 2002)
Back-up and Supportive Behavior	Back-up and supportive behavior refers to proactively seeking and providing task-related assistance, including the identification, remediation, and feedback on errors and near misses.	(Salas et al., 2005; Wilson, Salas, Priest, & Andrews, 2007)
Team Decision-Making	Team decision-making refers to the team's ability to determine goals, develop plans and strategies for task accomplishment, identifying contingencies, and updating/revising goals.	(Marks et al., 2001; Cannon-Bowers et al., 1995; Reader et al., 2009)

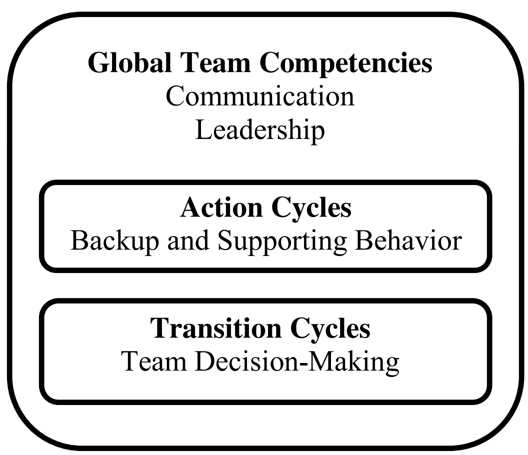

Figure 6. Relationship of ICU team performance dimensions in relation to action and transition cycles of team task accomplishment to guide measurement.

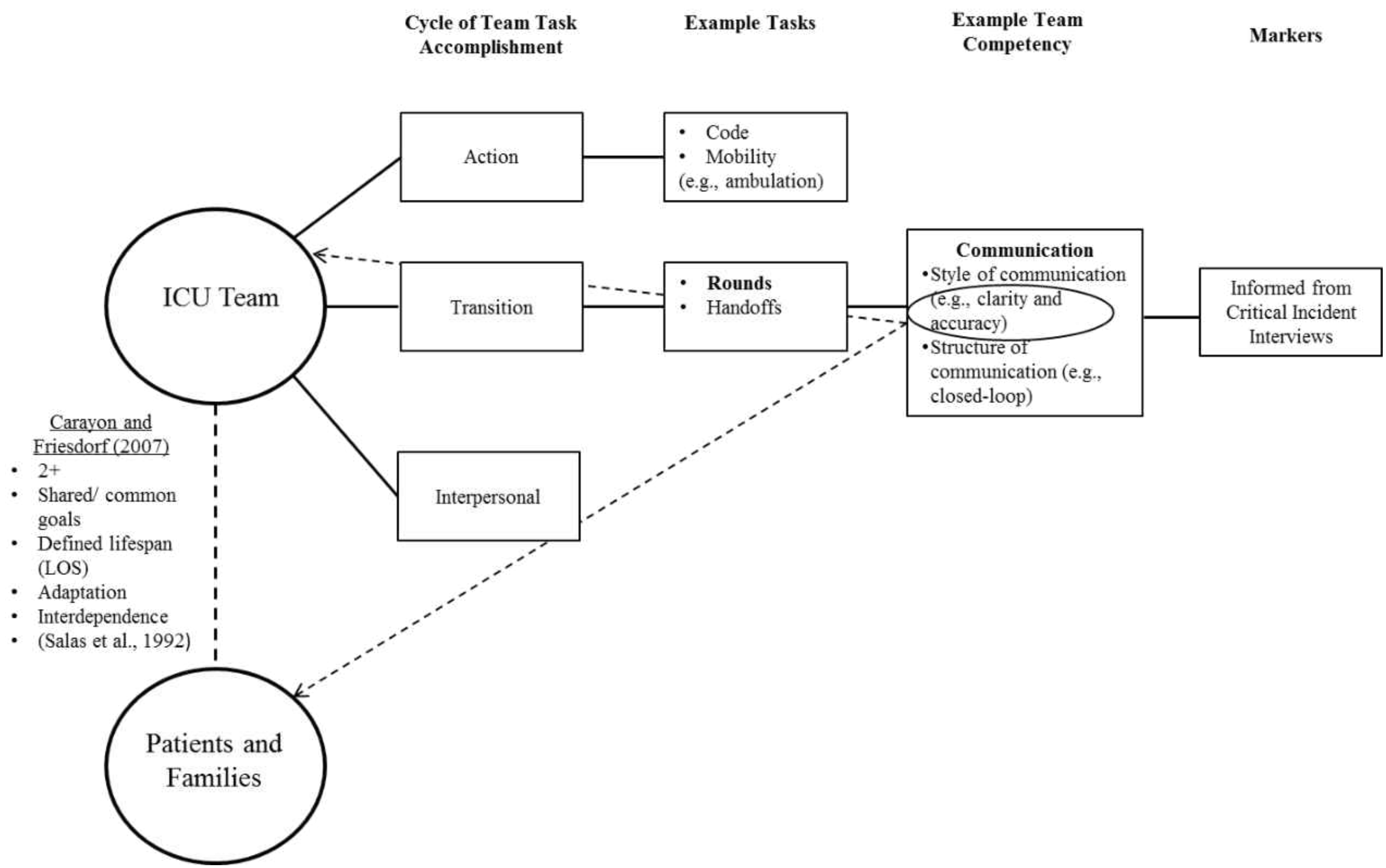

Figure 7. A framework for marker system development.

Identifying Behavioral Markers

The previous section demonstrated the theoretical and practical basis for selecting competencies to target for measurement. In addition to helping identify what should be targeted for measurement, subject matter expert guidance is also needed to provide a valid account of how these behaviors manifest. It is one thing to articulate a theoretical definition for a competency and another to generate explicit, observable indicators of performance (i.e., how the construct is operationalized). Therefore, another aspect of content analysis involved coding specific behavioral examples of ICU teamwork. As described earlier, participants were asked to recall positive and negative examples of teamwork using the critical incident technique. These behavioral examples were initially coded in relation to personnel and task type resulting in 192 examples of team performance. To illustrate, consider the following behavioral example from one interview:

> *So communication is very important. You should have a closed feedback loop. So when the leader assigns something to let's say the nurse, the leader should be hearing the order back from the nurse (e.g., "ok, I am going to grab the epinephrine").*

In this case, the participant was describing the importance of teamwork in relation to a code event between the individual leading a code and a supporting team member. Subsequently, behavioral examples were translated into markers that can be used to guide observation (i.e., action statements), as conveyed below:

> *Team member confirms they understand the directive and verbalizes their intent to execute the directive.*

This processes yielded 283 possible behavioral markers from the 192 examples of teamwork that were extracted from the critical incident interviews. In order to increase the accessibility and generalizability of the marker system, the next step of data reduction involved selecting behavioral markers that were task-generic (ostensible duplicates were also removed at this stage). This step reduced the overall repository of behavioral markers from 283 to 134.

As described earlier, four core dimensions of ICU teamwork were identified from interview findings, the ICU team-based literature, and science of teams. The pool of 134 possible behavioral markers were coded by the author and a human factors psychologist to qualitatively determine how well the markers 'fit' with the proposed dimensions of teamwork. Within each dimension, clear sub-dimensions emerged. For example, behavioral markers relating to *Backup and Supporting Behavior* 'loaded' onto three sub-dimensions: *Offering Support*, *Seeking Support*, and *Feedback*. Once all the behavioral markers were clustered, behavioral markers with similar content were combined and redrafted to reduce the overall pool of potential indicators. To illustrate, the markers "Team members articulate care goals and methods of task execution" and "Team members specify the course of planned clinical activities and methods for achieving care goals for each patient" were revised to the single marker "Team members specify methods for achieving care goals for each patient." This step of data reduction resulted in 87 markers.

The remaining 87 markers were then dichotomized into examples of 'good' and 'poor' performance as an assessment aid (ANTS; NOTSS). Additional tool refinement during this phase involved the author and a human factors psychologist generating negative examples of performance when needed. To illustrate, "Team leader provides assistance and feedback to residents when they execute unfamiliar tasks" is an example of a good teamwork behavior for the sub-dimension *Feedback*. In this case, a new marker was generated to reflect a negative indicator that raters can reference to guide measurement: "Sr. physician intervenes without explaining rationale to Jr. physician."

Refining the Marker System

Cognitive interviews were conducted with experienced ICU clinicians (two nurses and one attending). This approach, as described by the American Institutes for Research (2010) during the initial validation of the TeamSTEPPS-Teamwork Perceptions Questionnaire, helped identify items needing revision due to ambiguity or possible misinterpretations. Additionally, the author and a human factors psychologist practiced using the marker system during six handoffs and eight rounds. While this served as an initial rater calibration exercise (Chapter 5), marker system content was also refined during these observations. In total, efforts refining marker system content resulted in 75 behavioral markers indicative of good and poor teamwork.

A taxonomy of ICU teamwork developed throughout this process of marker identification and refinement is conveyed in Table 10. The ICU behavioral marker system developed for this study is presented in Appendix D. The structure of the marker system was informed by best practices from the healthcare literature (Chapter 3). The marker system incorporates descriptive information about the team task that might be of interest to researchers (type of task, duration, team size, team size variability, and team diversity). Ratings are made at the subdimension and dimension levels using a five-point behaviorally anchored scale. The endpoints and midpoint of the scale include behavioral anchors, while the intermediary points allow for subjective interpretation between each anchor. Raters can write notes during the course of the observation at the dimension and subdimension levels to guide measurement and facilitate feedback.

Table 10. Taxonomy of ICU teamwork.

Teamwork Dimension	Sub-Dimension	Example Behavioral Markers
Communication	Style	• Uses lay terms when discussing care plan with patients and/or family members. • Multiple speakers presenting information simultaneously (negative). • Volume is too low and pace is fast (negative).
	Content	• Appropriate communication protocols/tools are used/followed. • Big picture summaries are provided
	Closed-Loop	• Directive confirmed and intent to execute verbalized. • Receipt of communication acknowledged for both face-to-face and electronic communication.
Leadership	Delegation	• Roles and responsibilities delegated clearly • Expectations of task-work assignments are not established (negative).
	Norms	• Team leader acknowledges good work and provides positive reinforcement. • New team members introduce themselves to the clinical team.
Backup and Supportive Behavior	Offering Backup/ Support	• Reallocates work when a more critical task is presented. • Offers help throughout the shift.
	Seeking Backup/Support	• Immediately requests assistance during acute situation. • Recognizes when overloaded and engages appropriate resources
	Feedback	• Identifies errors/near misses and assists with remediation. • Sr. clinician intervenes without explaining rationale (negative).
Team Decision-Making	Planning and Establishing Goals	• Team members deliberately discuss, propose, and prioritize the planned course of patient care for each patient. • Anticipated outcomes of treatment activities are not identified (negative).
	Contingency Planning	• Identifies conditions or events that may alter treatment plans, including barriers and challenges that may impede progress. • Specifies alternative courses of action for treatment plans.
	Updating and Revising	• Identify any challenges encountered while executing care plans and emerging issues. • Relevant team members (including P/F) are informed of updates to care goals and pans, changing patient conditions, and following consults with inter-unit staff.

CHAPTER 5: METHODOLOGY

A multifaceted research approach must be utilized to provide a thorough evaluation of reliability and validity. Table 11 provides an overview of the methodological approach. This study provides an initial assessment of reliability and validity by focusing on a subset of teamwork competencies relevant to transition cycles of team task accomplishment.

Table 11. Overview of methodological approach.

Type of Reliability/Validity	Sources of Evidence
Content Validity	• ICU teamwork review • Critical incident interviews
Interrater Reliability	• Intraclass correlations (ICC) • Percent overall agreement
Generalizability Theory (Reliability)	• Provides reliability evidence by modelling systematic variance associated with rater, task, and random error effects.
Generalizability Theory (Validity)	• Provides validity evidence by modelling systematic variance associated with subdimension effects.

Observation Procedure

Data for this study were collected at a surgical ICU at Johns Hopkins Hospital. Two raters involved in the development of the behavioral marker system (the author and a human factors psychologist) rated the performance of teams as they completed handoffs (n=25) and rounds (n=25). Each rater scored six subdimensions for the present study: *Communication Style*, *Communication Content*, *Closed-loop Communication*, *Planning and Establishing Goals*, *Contingency Planning*, and *Updating and Revising Goals*.

Rater Training

Each rater was involved in the development of the marker system and had an understanding of the competencies targeted for measurement and associated markers of performance. Raters practiced using the system prior to conducting observations. Six handoffs and eight rounds were observed to this end. The practice observations only focused on scoring behaviors; characteristics of team size, size variability, and duration of each instance of teamwork were not documented. Following each instance of teamwork during practice observations, scores and discrepancies were methodically discussed for each competency that was rated. During these discussions, one rater would indicate how they scored a specific behavior and provide examples from the instance of teamwork to justify their rating. This was followed by the other rater explaining the score they gave to the same behavior and the rationale for that rating. At the completion of these practice observations, the raters understood why discrepancies took place and felt comfortable in how behaviors should be rated moving forward. Additionally, minor changes to the wording of markers were made to help prevent uncertainty.

Reliability and Validity Analyses

Content Validity

Findings from the systematic literature of ICU teamwork (Chapter 2) provided a first step towards identifying relevant behaviors. Critical incident interviews with subject matter experts (Chapter 4) formed the basis of what constructs to target in measurement and under what conditions they should be relevant (i.e., action or transition cycles of teamwork). For instance, *Team Decision-Making* will be more salient during transition cycles of team task accomplishment while *Backup and Supporting Behavior* will be more relevant during action phases of teamwork.

Interrater Reliability and Agreement

Interrater reliability signifies the extent to which scores made between two or more raters are proportional while interrater agreement specifies the degree to which scores made between two or more raters are exactly the same (Tinsley & Weiss, 1975). Intra correlations (ICC) were analyzed using *SPSS v.21* to index interrater reliability. A two-way, random effects model with absolute agreement was employed and both single and average measures are reported. The two-way random effects model is appropriate for generalizing to other raters randomly sampled from the population (Shrout & Fleiss, 1979). Absolute agreement was calculated as the percentage of exact matches in scores made between raters on the same target of measurement (i.e., one subdimension of teamwork). The percentages of scores within one point of each other are also reported.

Generalizability Theory: Reliability and Validity

Generalizability (G) theory was applied to further examine the reliability of the marker system and provide evidence of construct validity. This section begins by providing background on the approach and defining key terms then turns to a discussion of how G theory was employed for the present study.

Background

G theory was introduced by Lee Chronbach and colleagues in the volume *The Dependability of Behavioral Measurement* (Shavelson & Webb, 1991). It was developed as an extension of classical test theory (CTT) to allow researchers to examine the systematic effects of multiple sources of measurement error simultaneously through an analysis of variance (Brennan, 2001; Crossley, Davies, Humphris, & Jolly, 2002; Kraiger & Teachout, 1990; Shavelson & Webb, 1991).

Yet G theory is much more than an approach to model the error structure of a specific process of data collection. By applying an analysis of variance, researchers can also leverage G theory to provide construct related evidence (Kraiger and Teachout, 1990; Arthur, Woehr, & Maldegen, 2000). To illustrate, Kraiger and Teachout (1990) analyzed proficiency ratings in a four facet design that included different types of ratings forms, different rater sources (incumbents, peers, and supervisors), and items within the rating forms. The authors predicted variance associated with the individuals being rated would be the greatest to reflect individual differences in performance. They also expected a small person X forms interaction. The forms were developed to assess the same proficiency constructs, so a small variance component would suggest convergence over rating forms. Conversely, a larger persons X items within forms interaction would demonstrate that within a particular form, ratees are differentially ranked across items.

<u>Key Terms</u>

G theory defines sources of systematic variance as *facets* and levels of that facet as *conditions* (Kraiger & Teachout, 1990). For example, rater background may be a facet of generalization while behavioral scientist and clinician constitute conditions of that facet. Another facet may be the dimensions of the rating tool employed for a particular study. The variance associated with each facet is estimated through an analysis of variance to concurrently partition sources of systematic variance (Mathieu & Day, 1997).

As described by Cardinet et al. (2010), facets can be crossed or nested. A facet is *crossed* “when every level of one of the facets is combined with every level of the other in a data set” (p.13). For example, if every rater observers every team performing every task, the facets are crossed (see Table 12). Conversely, facets are nested when “each level of one [facet] is associated with one and only one level of the other” (p.13). For example, If Rater A only assessed teams when they completed Task 1 and Task 2 while Rater B only assessed teams when they completed Task 3 and Task 4, then Tasks would be nested within Raters. Because every team is still completing every task, however, teams are crossed with tasks and raters (see Table 13).

Table 12. Hypothetical research design for the crossing of Teams X Raters X Tasks.

Raters	Rater A				Rater B			
Tasks	Task 1	Task 2	Task 3	Task 4	Task 1	Task 2	Task 3	Task 4
Team 1								
Team 2								
…								
Team k								

Table 13. Hypothetical research design for the nesting of Tasks within Raters.

Raters	Rater A		Rater B	
Tasks	Task 1	Task 2	Task 3	Task 4
Team 1				
Team 2				
…				
Team k				

Facets are also treated as *fixed* or *random*. While fixed facets assume all possible conditions for a facet were selected for a study, random facets assume conditions were randomly selected from a population under investigation (Cardinet et al., 2010). For instance, subdimensions of the ICU marker system are considered to be random because they represent a sample of all the possible items that can be used to assess teamwork.

Application to Present Study

By specifying potential sources of systematic variance and the magnitude of that variance, G studies provide a powerful resource for providing evidence of reliability and validity. Four sources of systematic variance were analyzed for the present study: instances of teamwork (I), rater effects (R), subdimension effects of the marker system (S), and task effects (T). An overview of the study design is presented in Table 14.

It is first worth noting that instances of teamwork are nested within the task facet, with teams being operationally linked to the tasks they perform. Consider a team task such as daily rounds. For each patient, there will be an attending, fellow, resident, nurse, medical student, and other ancillary staff participating in the round. While attendings and fellows are typically assigned to manage all of the patients on a given unit, residents are assigned a subset of patients (one-half or one-third of the unit) and nurses are assigned an even smaller subset (typically one to two). Ancillary staff such as pharmacists, nutritionists, and physical therapists may be involved for a particular patient or every patient. This means that each round constitutes a different instance of teamwork. Similarly, a code event will involve a unique subset of team members. Typically, the fellow (and/or attending) as well as the bedside nurse for that patient will be present to respond to the code. While other unit personnel will also respond to the code (i.e., supporting behavior), the compositional makeup for a code team will vary from code event to code event. Thus, instances of teamwork are nested within the tasks—one team can only be observed for one task.

All instances of teamwork were observed by two raters for two tasks: rounds and handoffs. Although these tasks were not randomly sampled, there is precedent for convenience sampling from a larger population of a facet and treating that facet as random (Shavelson et al., 1990). Each rater assessed teamwork along six subdimensions of the marker system: *Communication Style*, *Communication Content*, *Closed-loop Communication*, *Planning and Establishing Goals*, *Contingency Planning*, and *Updating and Revising Goals*. These subdimensions were selected because each occurs during both rounds and handoffs, affording a crossed design. This facet is also treated as random for the present study.

Table 14. Overview of study design.

Task	Handoffs												Task	Rounds											
Raters	Rater 1						Rater 2						Raters	Rater 1						Rater 2					
Sub dimension													Sub dimension												
Instance H1	S	C	L	G	P	U	S	C	L	G	P	U	Instance R1	S	C	G	L	P	U	S	C	L	G	P	U
Instance H2	S	C	L	G	P	U	S	C	L	G	P	U	Instance R2	S	C	G	L	P	U	S	C	L	G	P	U
Instance H3	S	C	L	G	P	U	S	C	L	G	P	U	Instance R3	S	C	G	L	P	U	S	C	L	G	P	U
Instance H4	S	C	L	G	P	U	S	C	L	G	P	U	Instance R4	S	C	G	L	P	U	S	C	L	G	P	U
Instance H…25	S	C	L	G	P	U	S	C	L	G	P	U	Instance R…25	S	C	G	L	P	U	S	C	L	G	P	U

S = Communication Style
C = Communication Content
L = Closed-loop Communication
G = Planning and Establishing Goals
P = Contingency Planning
U = Updating and Revising

Figure 8 illustrates the variance components for the (I:T) X R X S design. Ideally, the study design would be fully crossed, meaning that every team was observed by every rater on every task and scored on every dimension of teamwork. A fully crossed design would afford the examination of all main effects and interaction effects of variance components identified in the study. Due to the nested design, however, not all variance components can be estimated due to confounding (Shavelson & Webb, 1991). In the current study, the instance main effect would be confounded by the instance X task interaction. This means that instance X task interaction cannot be estimated at all. Additionally, it is difficult to parse how the systematic variance changes depending on whether the data for rounds and handoffs are combined compared to when they are looked at separately. Therefore, the two additional G studies were conducted to account for findings that may be methodologically misleading due to nesting. The amalgamation of these studies will provide a sufficient opportunity explore evidence of the marker system's reliability and validity. G Study 1 accounted for the complexity of the nested design as depicted in Table 15. G studies 2 and 3 relied on the same measurement design, except the data from only handoffs (study 2) and rounds (study 3) were analyzed. The juxtaposition of variance components across tasks will afford a qualitative comparison of variance estimation (see Table 16).

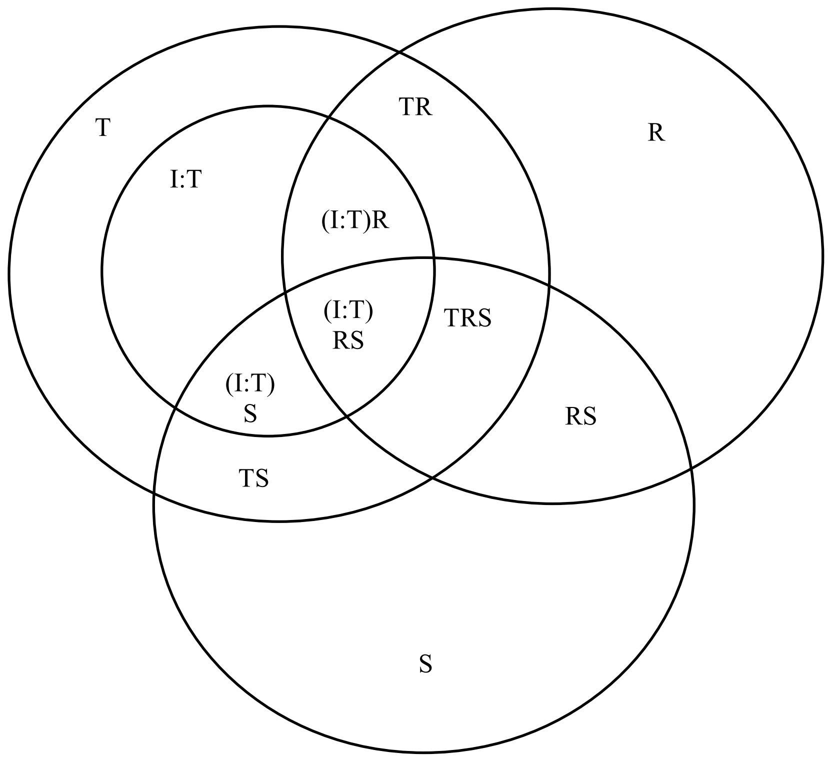

Figure 8. Variance components for ICU G study.
Note: T = Tasks; S = Subdimension; R = Raters; I:T = Instances of teamwork nested within tasks

Table 15. Sources of variability for ICU G study.

Source of Variation	**Description** (see: Arthur, Woehr, & Maldegen, 2000; Crossley et al., 2002; Crossley et al., 2007; Kraiger & Teachout, 1990)
T	• Systematic variances across instances, subdimensions, and raters.
I:T	• Systematic variance in instances of teamwork nested within tasks across subdimensions and raters.
S	• Systematic variance in subdimensions across instances and tasks.
R	• Systematic variance in ratings across subdimensions, instances of teamwork, and tasks.
RS	• Variance due to raters consistently scoring a particular teamwork subdimension differently.
T R	• Variance due to raters consistently scoring a particular task differently.
I:T R	• Variance due to raters consistently scoring teams within a particular task differently.
I:T S	• Variance due to teams within tasks preforming differently on subdimensions of teamwork.
I:T R S	• Residual error. Note: the T X R X S interaction is not distinguishable from the residual error term (Crossley et al., 2007).

Table 16. Example of variance components juxtaposed across tasks for G studies 2 and 3.

Source of Variation	Var(Handoffs)	Var(Rounds)
I		
R		
S		
IR		
RS		
IS		
IRS		

G Study Variance Analyses and Predictions

Data for the G studies were analyzed using *EduG v.6.1*, a software package developed specifically for G studies (Cardinet et al., 2010). Briefly, *EduG* enumerates sources of variance from a data set and calculates a G coefficient using a predetermined measurement design. The measurement design distinguishes differentiation facets (i.e., sources of desired variation) from instrumentation facets (i.e., sources of unwanted variance). G coefficients, like ICCs, represent the proportion of true score variance to observed score variance (Cardinet, Tourneur, & Allal, 1976).

The goal of the present study is to demonstrate that the marker system differentiates subdimensions for the tasks that were observed regardless of raters or instance of teamwork (i.e., generalizing variance in subdimensions to other instances of teamwork and raters). Thus, a large G coefficient serve as an indicator of construct validity because the marker system adequately discriminates among teamwork competencies.

Because G theory does not lend to hypothesis testing, no formal hypotheses can be made (Kraiger & Teachout, 1990). That said, a certain pattern of results is expected (Table 17). The variance associated with the Subdimension X Instance interaction should be the greatest, followed by the main effects for the subdimensions and instances. Variance associated with rater effects and the task main effect should be the smallest.

The subdimension X instance interaction should be the greatest source of variance because a large main effect for instances of teamwork would only provide evidence of systematic differences in the average ratings of teams (across raters and subdimensions). For example, if a particular team received a rating of '1' across all subdimensions, another team received ratings of '3' across all subdimensions, and yet another team received a rating of '5' across all subdimensions, the study would highlight differences in overall teamwork, but no variability in how subdimensions are scored. This would suggest that attributes and not differentially scored by raters. The same is true for the subdimension main effect: if a particular team received a rating of '1' for one subdimension, '2' for another subdimension, '3' for another subdimension, and so on for every instance that was observed, the study would highlight overall differences in how subdimensions are scored, but show no variability in how subdimensions are scored for a particular team.

Table 17. G study variance predictions.

Relative Variance Among G Study Variables
1. The I:T X S interaction in G Study 1 is expected to account for the most variance, indicating that teams within tasks differed on one subdimension of teamwork relative to another.
2. The R main effect, I:T X R interaction, S X R interactions in G Study 1 are expected to be small, indicating that raters consistently scored teams, subdimensions, and tasks.
3. The variance associated with the I:T main effect and the S main effect will be larger than variance associated with the R term, but not as large as variance associated with the I:T X S term.
4. For each team task in G studies 2 and 3, the I X S interaction 1 is expected to account for the most variance, indicating that teams differed on one subdimension of teamwork relative to another.
5. For each team task in G studies 2 and 3, The R main effect, I X R interaction, S X R interactions are expected to be small, indicating that raters consistently scored teams and subdimensions.
6. For each team task in G studies 2 and 3, The variance associated with the I main effect and the S main effect will be larger than variance associated with the R term, but not as large as variance associated with the I X S term.

G Study Sample Size Justification

The present study observed 25 teams for each task (n=50). Although G theory utilizes an analysis of variance, it is not a hypothesis testing model (Crossley et al., 2007). Therefore, sample size cannot be determined by estimating the number of participants needed to observe a small, medium, or large effect. In fact, there is no convention for determining an appropriate sample size for a G study. For this reason, observing 25 teams per task is methodologically defensible based on previous research efforts (Table 18).

Table 18. Summary of G studies and associated sample sizes.
Note: 'n' refers to the sample size of the object of measurement, not the levels of a facet of the reported studies.

Study	Description
(Crossley et al., 2011)	• n=85 • Five facets examined • Study duration: two years
(Weller et al., 2011)	• n=40 (teams) • Three facets examined
(Kraiger & Teachout, 1990)	• n=256 • Four facets examined
(Shavelson et al., 1990)	• n=150 • Seven facets examined

(Mathieu & Day, 1997)	• Study 1
	  ○ n=16
	  ○ Three facets examined
	• Study 2
	  ○ n=12
	  ○ Four facets examined

Ethics Approval

Data for the interrater reliability assessment and G studies were collected by the author as a Senior Research Program Coordinator II for the Johns Hopkins University, School of

Medicine. Approval for conducting this study was granted by the Institutional Review Boards at Johns Hopkins and the University of Central Florida (see Appendix E and F).

CHAPTER 6: RESULTS

Descriptive information about each team task is presented in Table 19. On average, rounds and handoffs lasted a similar duration. Surprisingly, a large percentage of instances of teamwork observed for handoffs were multidisciplinary. Team size was more stable for handoffs (56%) than rounds (36%), but the variability on both tasks was not expected.

Table 19. Characteristics of team tasks.

	Handoffs	**Rounds**
Average Team Size	2.84	10+
Max Team Size	5	N/A
Min Team Size	2	10+
Size Variability (Stable)	14 (56%)	9 (36%); 1 not documented
Composition (Multidisciplinary)	11 (44%)	25 (100%)
Average Length	13:47	13:38; 1 not documented
Max Length	25:35	29:36
Min Length	3:58	4:31

Rounds and Handoffs

Table 20 provides means and standard deviations for each teamwork subdimension when the data for handoffs and rounds are combined. Mean ratings for *Communication Style* were the lowest among both raters and also had the largest standard deviation. Mean ratings for *Updating and Revising Goals* were the largest among both raters, with the smallest standard deviation.

Table 20. Mean ratings of teamwork for handoffs and rounds.

Aspect of Teamwork	**Rater 1**	**Rater 2**
Overall (n=50)	4.21 (.90)	4.32 (.86)
Communication Style	3.58 (1.09)	3.68 (.98)
Communication Content	4.36 (.66)	4.52 (.70)
Closed-Loop Communication	4.3 (.74)	4.4 (.70)
Planning/ Establishing Goals	4.24 (.91)	4.46 (.79)
Contingency Planning	4.12 (.94)	4.18 (.96)
Updating/ Revising Goals	4.7 (.65)	4.68 (.65)

Table 21 presents intraclass correlations (ICCs) for each subdimension across rounds and handoffs. Both single measures and average measures are reported along with 95% confidence intervals. For single measures, all reliability calculations were consider good with the exception of *Contingency Planning*, which was fair. If average measures are considered, all reliability calculations are considered excellent, with the exception of *Contingency Planning*, which is considered good.

Table 21. ICCs for rounds and handoffs.

Aspect of Teamwork	**Single Measures***	**95% Confidence Interval**	**Average Measures***	**95% Confidence Interval**	**Percent Agreement**	**Within 1**
Overall	.68[c**]	.61 to .73	.81[d]	.76 to .85	.65	.95
Communication Style	.69[c]	.52 to .81	.82[d]	.68 to .90	.52	.94
Communication Content	.62[c]	.42 to .77	.77[d]	.59 to .87	.64	1.00
Closed-Loop Communication	.63[c]	.43 to .77	.76[d]	.61 to .87	.68	.98
Planning/ Establishing Goals	.61[c]	.40 to .76	.76[d]	.58 to .86	.66	.92
Contingency Planning	.52[b]	.28 to .70	.69[c]	.44 to .82	.64	.84
Updating/ Revising Goals	.74[c]	.58 to .84	.85[d]	.73 to .92	.78	1.00

*ICC model: 2-way random components, absolute agreement (see Shrout & Fleiss, 1979).
**Corresponding levels of practical, substantive, or clinical significance (see Cicchetti, 1994, p. 286)
a. Poor: <.40
b. Fair: .40 -.59
c. Good: .60 -.74
d. Excellent: ≥ .75

The analysis of variance attributions for each of the four main effects and associated interactions are presented in Table 22 (G Study 1). The pattern of results supports the expectation that the marker system differentiates among subdimensions. To illustrate, both the subdimension main effect (S) and the subdimension X instance interaction (IS:T) accounted for 43.5% of the total variance. Whereas the subdimension main effect evidences the marker system differentiates subdimensions scores (averaged over instances for each team task and raters), the subdimension X instance interaction demonstrates that for each instance of teamwork, subdimensions were differentially scored. Additionally, the instance main effect (I:T) accounted for 19.8% of the variance. This indicates that scores of teamwork (averaged over subdimensions and raters) were systematically different as well.

In contrast, the main effects associated with rater (R) and task (T) effects were very low, accounting for 7.5% of the total variance. This means there were minimal systematic differences in how raters scored subdimensions or tasks. The task X subdimension (TS) interaction indicates that average ratings of subdimensions were generally stable across each task (5% of estimated variance). Yet because teams are nested within each task, it is difficult to understand the true effect of the task effect. The remaining terms (IRS:T and TRS) constitute residual error variance (29.2%). Last, it is worth noting the small negative variance associated with some of the variables. In reality, variance cannot be negative, but a small negative variance is not uncommon when applying G theory (e.g., Shavelson et al. 1990, Cardinet et al., 2010). A small value, as is the case in the present analysis, is treated as null. A large negative variance (i.e., relative to other estimated variance components) would suggest some sort of model misspecification (Shavelson et al., 1990).

Table 22. G study 1: Analysis of variance for rounds and handoffs.

Source	SS	df	MS	Estimated Variance	%	SE
I:T	133.00667	48	2.77097	0.16161	19.8	0.04688
T	3.37500	1	3.37500	-0.00385	0.0	0.01062
R	1.60167	1	1.60167	0.00544	0.7	0.00439
S	64.20833	5	12.84167	0.10123	12.4	0.07026
IR:T	15.60667	48	0.32514	0.01447	1.8	0.01142
IS:T	178.75333	240	0.74481	0.25325	31.1	0.03555
TR	0.04167	1	0.04167	-0.00191	0.0	0.00100
TS	13.95500	5	2.79100	0.04086	5.0	0.02998
RS	0.84833	5	0.16967	-0.00144	0.0	0.00316
IRS:T	57.19333	240	0.23831	0.23831	29.2	0.02166
TRS	1.20833	5	0.24167	0.00013	0.0	0.00524
Total	469.79833	599			100%	

Table 23 presents differentiation facets in relation to error variance and also presents G coefficients. The sources of differentiation (i.e., desired variance) are listed in the second column and error variance for relative and absolute measurement considerations is presented in the remaining columns. The current measurement design helps answer the question: does the behavioral marker system differentiate subdimensions for the two tasks that were observed regardless of the instance of teamwork or who the rater is?

It is worth acknowledging the inclusion of the task term may seem inappropriate as a source of differentiation. The rational for its inclusion is that the subdimension X task interaction is not necessarily the product of measurement error. For example, such a finding might inform leadership of systematic differences in how teams are performing on certain attributes depending on what task they are executing. Conversely, error variance associated with particular instances of teamwork or raters may cloud systematic variance associated with the subdimensions. In the present case, however, there would be concern if the task main effect (T) contributed to a large amount of the differentiation variance. Such a finding would indicate that task effects alone (independent of subdimensions) are accounting for desired variance. This did not occur in the present study; task variance does not contribute to any differentiation variance.

Examining Table 23 below, the largest component of error variance is the instance X subdemension interaction (IS:T), which accounts for 41.5% of all error variance for absolute decisions. It is encouraging, however, that the differentiation variance is over five times larger than the error variance for both relative and absolute decisions. Additionally, the G coefficients for both relative (.87) and absolute (.85) decisions exceed standards for good reliability (.80) (Cardinet et al., 2010).

Table 23. Error variance and G coefficients for handoffs and rounds.

Source of variance	**Differ-entiation variance**	**Source of variance**	**Relative error variance**	**% relative**	**Absolute error variance**	**% absolute**
	.....	I:T	0.00646	29.8	0.00646	26.5
T	(0.00000)		.....		.....	
	.....	R	.....		0.00272	11.1
S	0.10123		.....		.....	
	.....	IR:T	0.00029	1.3	0.00029	1.2
	.....	IS:T	0.01013	46.6	0.01013	41.5
	.....	TR	(0.00000)	0.0	(0.00000)	0.0
TS	0.04086		.....		.....	
	.....	RS	(0.00000)	0.0	(0.00000)	0.0
	.....	IRS:T	0.00477	21.9	0.00477	19.5
	.....	TRS	0.00007	0.3	0.00007	0.3
Sum of variances	0.14208		0.02172	100%	0.02444	100%
Standard deviation	0.37694		Relative SE: 0.14737		Absolute SE: 0.15632	
Coef_G relative	0.87					
Coef_G absolute	0.85					

Handoffs

Table 24 reports means and standard deviations for each teamwork subdimension for handoffs that were observed. Mean ratings for *Communication Style* were the lowest among both raters and also had the largest standard deviation. Mean ratings for *Updating and Revising Goals* were the largest among both raters, with the smallest standard deviation. Both of these findings are consistent with the results reported for handoffs and rounds.

Table 24. Mean ratings of teamwork for handoffs.

Aspect of Teamwork	Rater 1	Rater 2
Overall (n=25)	4.13 (.95)	4.25 (.89)
Communication Style	3.6 (1.19)	3.72 (1.06)
Communication Content	4.32 (.63)	4.48 (.65)
Closed-Loop Communication	4.32 (.85)	4.56 (.71)
Planning/ Establishing Goals	3.92 (.95)	4.12 (.88)
Contingency Planning	3.84 (.90)	4.00 (1.00)
Updating/ Revising Goals	4.8 (.57)	4.64 (.64)

Table 25 presents ICCs for each subdimension for handoffs. Both single measures and average measures are provided along with 95% confidence intervals. The pattern of results when examining only handoffs differs from the results reported when both rounds and handoffs are considered. Only three variables indicated good reliability when single measures are considered (*Communication Style*, *Closed-loop Communication*, and *Updating and Revising Goals*). These competencies also demonstrated excellent reliability when average measures are considered.

Table 25. ICCs for handoffs.

Aspect of Teamwork	Single Measures*	95% Confidence Interval	Average Measures*	95% Confidence Interval	Percent Agreement	Within 1
Overall	.64[c**]	.53 to 0.73	.78[d]	.7 to .84	.63	.92
Communication Style	.64[c]	.33 to 0.82	.78[d]	.49 to .90	.44	.88
Communication Content	.52[b]	.17 to 0.75	.68[c]	.30 to .86	.60	1
Closed-Loop Communication	.69[c]	.40 to 0.85	.81[d]	.58 to .92	.72	.96
Planning/ Establishing Goals	.55[b]	.22 to 0.77	.71[c]	.36 to .87	.60	.88
Contingency Planning	.47[b]	.09 to 0.72	.64[c]	.17 to .84	.64	.80
Updating/ Revising Goals	.68[c]	.41 to .85	.81[d]	.58 to .92	.76	1

*ICC model: 2-way random components, absolute agreement (see Shrout & Fleiss, 1979).
**Corresponding levels of practical, substantive, or clinical significance (see Cicchetti, 1994, p. 286)
a. Poor: <.40
b. Fair: .40 -.59
c. Good: .60 -.74
d. Excellent: ≥ .75

The analysis of variance attributions for each of the three main effects and associated interactions are presented in Table 26 (G Study 2). The pattern of results is very similar to the results reported for both handoffs and rounds, with variance associated with the subdimension facet accounting for 44.5% of the total variance. Interestingly, there was a small amount of variance associated with the instance X rater interaction (3.5%). This indicates there may have been a small percentage of cases where the average scores of instances of teamwork (i.e., the average of each subdimension over a single instance of teamwork) were systematically different between raters. Like the results reported for handoffs and rounds, this G study includes a relatively large term for unexplained error variance.

Table 26. G study 2: Analysis of variance for handoffs.

Source	SS	df	MS	Estimated Variance	%	SE
I	75.12000	24	3.13000	0.18094	20.6	0.07363
R	1.08000	1	1.08000	0.00428	0.5	0.00601
S	38.50667	5	7.70133	0.13883	15.8	0.08240
IR	10.92000	24	0.45500	0.03039	3.5	0.02182
IS	93.16000	120	0.77633	0.25183	28.7	0.05268
RS	1.28000	5	0.25600	-0.00067	0.0	0.00565
IRS	32.72000	120	0.27267	0.27267	31.0	0.03491
Total	252.78667	299			100%	

The differentiation facet in relation to error variance is presented in Table 27 along with G coefficients for handoffs. This measurement design helps inform whether the behavioral marker system differentiates subdimensions regardless of the instance of teamwork that was observed or who the rater was when only handoffs are considered. As demonstrated in Table 27, the behavioral marker appears to adequately differentiate among constructs during handoffs. The differentiation variance is five times larger than both the relative and absolute error variance. The G coefficients for both relative (.90) and absolute (.84) decisions indicate the behavioral marker system reliability distinguishes teamwork competencies.

Table 27. Error variance and G coefficients for handoffs.

Source of variance	**Differ-entiation variance**	**Source of variance**	**Relative error variance**	**% relative**	**Absolute error variance**	**% absolute**
	.....	I	.....		0.00724	28.4
	.....	R	.....		0.00214	8.4
S	0.13883		.....		.....	
	.....	IR	.....		0.00061	2.4
	.....	IS	0.01007	64.9	0.01007	39.5
	.....	RS	(0.00000)	0.0	(0.00000)	0.0
	.....	IRS	0.00545	35.1	0.00545	21.4
Sum of variances	0.13883		0.01553	100%	0.02551	100%
Standard deviation	0.37260		Relative SE: 0.12461		Absolute SE: 0.15972	
Coef_G relative	0.90					
Coef_G absolute	0.84					

Rounds

Table 28 provides means and standard deviations for each teamwork subdimension for rounds that were observed. Mean ratings for *Communication Style* were the lowest among both raters and also had the largest standard deviation. Mean ratings for *Updating and Revising Goals* were the largest among both raters, though the standard deviation for *Closed-Loop Communication* was the smallest. These findings are similar to the results reported for handoffs and rounds/handoffs.

Table 28. Mean ratings of teamwork for rounds.

Aspect of Teamwork	Rater 1	Rater 2
Overall (n=25)	4.23 (.93)	4.40 (.88)
Communication Style	3.56 (1.00)	3.64 (.91)
Communication Content	4.4 (.71)	4.6(.77)
Closed-Loop Communication	4.28 (.61)	4.24 (.66)
Planning/ Establishing Goals	4.56 (.77)	4.8 (.5)
Contingency Planning	4.4 (.91)	4.36 (.91)
Updating/ Revising Goals	4.6 (.71)	4.72 (.69)

Table 29 presents ICCs for each subdimension for rounds, with single and average measures differing on some competencies from those reported for handoffs. The reliability calculations for *Communication Style* and *Updating and Revising Goals* were excellent, *Communication Content* was good, and the rest of the subdimensions were fair when single measures were considered. Additionally, the ICCs for *Communication Content* were greater for rounds than they were for handoffs, while *Closed-loop Communication* was worse. *Communication Style*, *Communication Content*, and *Updating and Revising Goals* demonstrated excellent reliability when average measures are considered and *Closed-loop Communication*, *Planning and Establishing Goals*, and *Contingency Planning* exhibited good reliability.

Table 29. ICCs for rounds.

Aspect of Teamwork	Single Measures*	95% Confidence Interval	Average Measures*	95% Confidence Interval	Percent Agreement	Within 1
Overall	.70[c**]	.62 to .77	.83[d]	.77 to .87	.67	.95
Communication Style	.78[d]	.56 to .90	.88[d]	.72 to .95	.6	1
Communication Content	.71[c]	.45 to .86	.83[d]	.62 to .93	.68	1
Closed-Loop Communication	.55[b]	.20 to .78	.71[c]	.34 to .87	.64	1
Planning/ Establishing Goals	.55[b]	.21 to .77	.71[c]	.35 to .87	.72	.96
Contingency Planning	.53[b]	.18 to .76	.70[c]	0.30 to .87	.64	.88
Updating/ Revising Goals	.79[d]	.59 to .90	.89[d]	.74 to .95	.8	1

*ICC model: 2-way random components, absolute agreement (see Shrout & Fleiss, 1979).
**Corresponding levels of practical, substantive, or clinical significance (see Cicchetti, 1994, p. 286)
a. Poor: <.40
b. Fair: .40 -.59
c. Good: .60 -.74
d. Excellent: ≥ .75

The analysis of variance attributions for the three main effects and associated interactions are reported in Table 30 (G Study 3). A similar pattern of results is demonstrated as the examination of handoffs and rounds/handoffs. Variance associated with the subdimension main effect and subdimension X instance interaction account for 53.4% of the total variance. Variance associated with the instances main effect accounts for 19% of the total variance. Only 0.4% of the total variance is associated with rater effects. Last, a large proportion of the variance is attributable to unexplained error variance (27.2%).

Table 30. G study 3: Analysis of variance for rounds.

Source	SS	df	MS	Estimated Variance	%	SE
I	57.88667	24	2.41194	0.14228	19.0	0.05649
R	0.56333	1	0.56333	0.00278	0.4	0.00314
S	39.65667	5	7.93133	0.14533	19.4	0.08483
IR	4.68667	24	0.19528	-0.00144	0.0	0.01002
IS	85.59333	120	0.71328	0.25467	34.0	0.04749
RS	0.77667	5	0.15533	-0.00194	0.0	0.00348
IRS	24.47333	120	0.20394	0.20394	27.2	0.02611
Total	213.63667	299			100%	

Table 31 presents the amount of differentiation variance in relation to error variance for rounds along with G coefficients. The pattern of results is similar to the results reported for handoffs. The largest contributor to error variance is the instance X subdimension interaction, which is slightly higher for rounds than handoffs. The differentiation variance is nearly 10 times greater than relative error variance and nearly seven times greater than the absolute error variance. The G coefficients for both relative (.91) and absolute (.87) decisions are also favorable. These findings support the expectation that the behavioral marker system reliability distinguishes teamwork competencies for rounds.

Table 31. Error variance and G coefficients for rounds.

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
	.....	I	.....		0.00569	26.7
	.....	R	.....		0.00139	6.5
S	0.14533		.....		.....	
	.....	IR	.....		(0.00000)	0.0
	.....	IS	0.01019	71.4	0.01019	47.7
	.....	RS	(0.00000)	0.0	(0.00000)	0.0
	.....	IRS	0.00408	28.6	0.00408	19.1
Sum of variances	0.14533		0.01427	100%	0.02135	100%
Standard deviation	0.38123		Relative SE: 0.11944		Absolute SE: 0.14610	
Coef_G relative	0.91					
Coef_G absolute	0.87					

CHAPTER 7: DISCUSSION

The goal of this study was to develop and test the validity of a behavioral marker system to assess ICU team performance. The first task involved identifying which teamwork competencies were most relevant to ICU teams and when they were important to lay the foundation for measurement. To this end, two systematic literature reviews and critical incident interviews with ICU clinicians were conducted. This effort resulted in a behavioral marker system that was both theoretically-based and context relevant.

Unquestionably, important aspects of teamwork are absent from the behavioral marker system. The development process involved the delicate balance of ensuring the most relevant aspects of ICU teamwork were incorporated in the tool while ensuring raters would not be overloaded with content to attend to during observations. For instance, Reader and colleagues (2011) provided an insightful account of team leadership behaviors in critical care. Guided by prominent theory in team leadership (e.g., Burke et al., 2006; Hackman, 2002), the authors identified 78 functional and developmental behaviors. While such a comprehensive inclusion and conceptualization of information would be valuable for a marker system dedicated specifically to ICU team leadership, it would not be practical to provide such a robust account of a single competency for a general tool.

The next phase of this study was dedicated to testing the reliability and construct validity of the marker system. Findings from the previous chapter provided initial evidence the behavioral marker system reliably differentiates among six teamwork competencies: *Communication Style*, *Communication Content*, *Closed-loop Communication*, *Planning and Establishing Goals*, *Contingency Planning*, and *Updating and Revising Goals*. The communication attributes represent a global teamwork competency while the team decision-making attributes represent a transition-oriented competency (Figure 6). Therefore, these data and conclusions do not generalize to action-oriented team tasks or other marker system content.

G coefficients, like ICCs, represent the amount of variance in observed scores attributable to true score variance (Mathieu & Day, 1997). In each of the G studies that were conducted, G coefficients exceeded conventional standards for both relative and absolute decisions (Cardinet et al., 2010). This means the marker system can be used to understand relative differences between teams (e.g., there is a 2 point difference in team A's quality of *Communication Style* relative to team B) as well as make absolute distinctions (e.g., team A scored a 5 on *Communication Style* while team B scored a 3).

The confluence of evidence presented in the previous chapter supports the expectation that raters involved in the development of the marker system can differentiate among six different competencies for rounds and handoffs. While these results are encouraging, they should also be interpreted with caution prior to implementing the marker system further. Different reliability indices provided varying levels of confidence in rater reliability and agreement. For example, the G studies indicated there were no systematic differences in how instances of teamwork or subdimensions were scored between raters by task. The ICCs and percent of absolute agreement, however, were not as encouraging. Because this study applied a psychometric approach, these and other considerations for improving the tool for future research and development can be addressed, as described in the remainder of this chapter.

Comparing Different Sources of Evidence

Generalizability Theory

The advantage of applying G theory is to model good and bad sources of systematic variance in ratings. By using this approach, it is possible to detect and enumerate variance that is the result of instances of teamwork, raters, subdimensions, and tasks in a single empirical design. Table 32 compares the percentage of each source of variation for when the data for rounds and handoffs are combined and when they are looked at separately for ease of interpretation. Variance due to rater effects (R) was low across each G study (0.7% overall, 0.5% for handoffs, and 0.4% for rounds). This indicates there were minimal systematic differences in how instances of teamwork were scored by raters. There were also no systematic differences in how raters scored a particular subdimension (RS) for either task. A small amount of variance for handoffs is attributed to the rater X instance interaction (3.5%), but not for rounds. This means there may have been a subset of handoffs where one rater systematically scored instances (averaged over subdimensions) higher than the other.

The subdimension X instance interaction accounted for a large proportion of the variance for each G study, with rounds explaining slightly more variance (34%) than handoffs (28.7%). This finding supports the expectation that for each instance of teamwork, subdimensions are differentially scored by raters. The subdimension main effect (20.6%) is slightly smaller than the instance main effect (15.8%) for handoffs while the subdimension main effect (19.4%) and instance main effect (19%) were nearly identical for rounds. Finally, a substantial amount of variance was attributed to residual error (29.2% overall), with the error term being slightly higher for handoffs (31%) than rounds (27.2%).

Table 32. Comparison of sources of systematic variation across G studies.

Source of Variation	% Var(Overall)	% Var(Handoffs)	% Var(Rounds)
I:T	19.8	…	…
T	0.0	…	…
I	…	20.6	19.0
R	0.7	0.5	0.4
S	12.4	15.8	19.4
ST	5.0	…	…
RT	0.0	…	…
IR:T; IR	1.8	3.5	0.0
RS	0.0	0.0	0.0
IS:T; IS	31.1	28.7	34.0
IRS:T; IRS	29.2	31.0	27.2
G Coefficient Relative	0.87	0.90	0.91
G Coefficient Absolute	0.85	0.84	0.87

Despite illustrating a pattern of results consistent with expectations, there was a large residual error variance associated with each G study. These findings converge to suggest approximately 30% of variance associated with each study constituted residual error. This residual error, however, did not attenuate the G coefficient. That said, the underlying cause of the residual error warrants further discussion. First, the experience of team members that were observed during instances of teamwork could have influenced ratings. For example, case presentations during rounds are generally given by a medical student or a resident, but could also be given by fellows and the attending. Although information about who was giving the presentation was not captured, this variable may have influenced ratings of certain teamwork dimensions. Similarly, the unit also employs traveling nurses to fill scheduling voids. These nurses may not have as much familiarity with the structured protocols for conducting handoffs, thereby influencing ratings of teamwork dimensions (e.g., reviews of patient info may not be presented by system).

Patient complexity presents another possible contributor to ratings of teamwork that was not defined by the G studies. Complex patients may require more resource and contingency planning, opportunities to teach (for rounds), information to transfer between shifts (handoffs), and greater multidisciplinary input. Objective data such as the patient's length of stay, ventilator status, and number and types of medications could be collected in future studies to determine whether patient conditions predict teamwork and team outcomes (or influence rater reliability).

Another variable that may have influenced ratings of teamwork was a newly introduced nursing huddle that occurred at the beginning of each shift (0700). During this huddle, oncoming nurses are paged to meet in a central location and learn about issues relating to unit census for the day. These huddles lasted approximately 1-2 minutes and interrupted seven handoffs that were observed. These huddles may have disrupted the flow of the handoff and thereby influenced ratings of behaviors.

It is also worth considering how the variance attributions would look if ratings were made at the dimension level. Such a design would involve two nested terms (instances nested within tasks and subdimensions nested within dimensions). Conceptually, a study design that includes additional facets should limit the amount of residual error because there is more variance in the model that can be explained. This design could also inform the extent to which there are differences in how ratings are made at the dimension level (good variance) or whether judgments of behavior are clouded by newly introduced rater error (i.e., bad variance associated with the rater). As additional opportunities for data collection become available, more facets can be defined for future analyses to better understand the extent to which residual error pervades the variance in ratings. For example, different attendings (i.e., leaders) may influence the quality of teamwork during rounds. Exploring the systematic variance of instances of teamwork (nested within attending) can inform whether the marker system generalizes across different attendings in a particular unit as well as whether certain attendings performed better (or worse) on subdimensions or instances of teamwork relative to other attendings.

Intraclass Correlations

Overall, the ICCs were good for the study of rounds and handoffs, handoffs, and rounds when single measures are considered. The average measures index was always higher than single measures, but this result is expected because more measurements are evaluated. Despite the low ICCs for single measures, the results are comparable to the NOTSS marker system developed for the assessment of surgeons' non-technical skills. Yule and colleagues (2008) reported the following ICCs for dimension ratings of the NOTSS system: situational awareness (.29), decision-making (.60), task management (.39), leadership (.66), and communication and teamwork (.63). Average measure calculations, however, were much greater: situational awareness (.95), decision-making (.99), task management (.97), leadership (.99), and communication and teamwork (.99).

Teamwork in healthcare and within the ICU is complex, making the judgment of team behaviors challenging as well. Marker systems attempt to alleviate some of these challenges by providing additional structure to assessment. Certainly, areas of the ICU marker system warrant further consideration during refinement or should be emphasized in future rater training. Yet some of the challenges are intrinsic to the rating context. To illustrate, a single statement from a clinician in the real world could involve behaviors related to *Updating and Revising Goals* (e.g., the patient did respond to a certain treatment), *Planning and Establishing Goals* (e.g., consults with outside services and/or additional tests are suggested), and *Contingency Planning* (e.g., there are no signs of active bleeding, but that is a situation in need of monitoring). Capturing all of this information is a difficult undertaking for raters, especially when the behaviors occur in rapid succession.

Another explanation for the low ICCs reported from the previous chapter may be due to a range restriction of attribute scoring. In the present study, the extreme negative end of the scale was not used for the instances of teamwork that were observed while the score of ‘2’ was only given 24 times of 300 possible ratings. Even though the G studies provided evidence that raters scored each subdimension differently, teams may generally be homogenous on each teamwork competency. The ICC indexes the variance of instances of teamwork that were observed with the variance between raters; the small variance of how instances were scored may be contributing to unfavorable ICCs by deflating the true score variance in the ICC equation and amplifying rater discrepancies (Gaba et al., 1998; Tinsley & Weiss, 1975). In the ICU where observations were conducted, structured protocols were in place for both handoffs and rounds. These systems are mechanisms that improve teamwork (see Chapter 2) and may be contributing to the small variance because all teams are generally performing well. Future research would benefit from using the marker system in a different unit or hospital to observe a more heterogeneous sample (i.e., there is a need to assess the viability of this tool in other research contexts to evaluate its generalizability in places with different norms and organizational policies). Further, testing the psychometric properties of the marker system in another research context affords the examination of additional facets in a G study, which may explain a greater percentage of the residual error reported in the previous chapter.

In contrast to the analysis of variance reported for each G study, which showed minimal differences in how subdimensions were rated by task, examination of ICCs paints a different picture. Figures 9 and 10 present the ICCs across rounds and handoffs (overall), handoffs, and rounds for both single and averages measures respectively. *Communication Style* and *Communication Content* showed greater inconsistency during handoffs than rounds. *Updating and Revising Goals* showed greater inconsistency during handoffs than rounds as well, though the ICCs ranged from good (for single measures) to excellent (for average measures) in both cases. Last, *Closed-loop Communication* showed greater inconsistency in rounds than handoffs. The ICU behavioral marker system developed for this study ultimately aims to generalize across multiple task types. This finding may suggest more context specific guidance is needed for the markers themselves or that certain competencies should be given greater emphasis in rater training.

It is also worth noting that the ICC for *Contingency Planning* was consistently low across each analysis, never surpassing a fair rating when single measures are considered. This finding is consistent with the percent of overall agreement for this competency as well. This particular competency may have been low for a variety of reasons. First, as described earlier, homogeneity of variance may be contributing to the low ICCs; the small variability in true scores between instances of teamwork may be magnifying the inconsistencies in scoring attributions between raters. Second, it may have been difficult for raters to capture and evaluate this competency. Findings from the critical incident and cognitive interviews provided support for *Contingency Planning* being a unique competency for ICU teams. It differs from *Planning and Establishing Goals* in that alternative treatment courses are identified and rationalized. That said, it may have been difficult for raters to (1) separate these distinctions during actual observations and (2) evaluate whether *Contingency Planning* was actually thorough (e.g., quality and quantity). Future rater training and guidance would benefit from providing explicit examples of what constitutes good and poor performance along the entire spectrum of the rating scale for each teamwork competency.

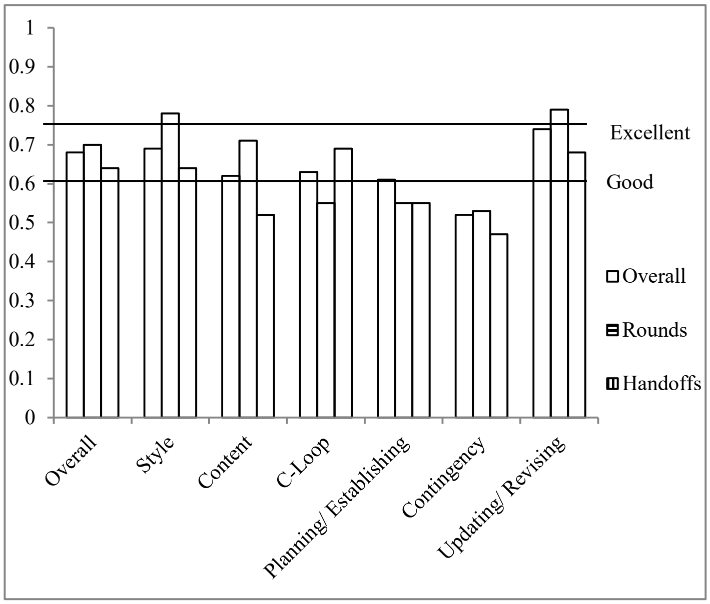

Figure 9. Comparison of single measure ICCs.

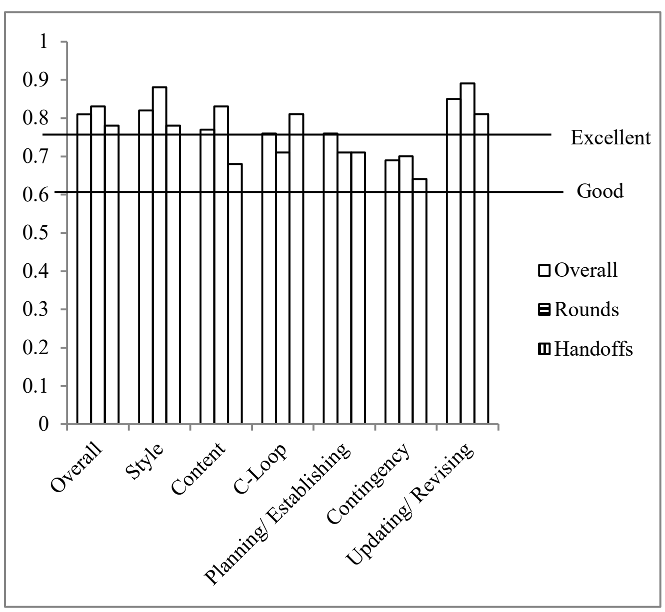

Figure 10. Comparison of average measure ICCs.

A final area contributing to the low ICCs may simply be rater biases. One error that raters may be particularly susceptible to in this performance context is the contrast effect. The contrast effect occurs when raters compare a current instance of teamwork to a previous instance of teamwork when making performance valuations, rather than relying on behavioral markers and the behaviorally anchored scale (Feldman et al., 2012). To illustrate, one team may have been considerably exceptional and received ratings of five for each subdimension. If raters gave the subsequent team ratings of four because they did not perform as well as the previous team even though they performed to the standards congruent with a rating of five, then the contrast took place. Thus, error is introduced to ratings, leaving scores artificially deflated or inflated (and inconsistent) across measurements. The contrast effect may have been especially applicable to rounds, where raters had the opportunity to assess as many instances of teamwork as there were beds in the unit in a single day.

Percent of Overall Agreement

Although not a generally accepted index of rater reliability (Hallgren, 2012), percent of overall agreement was calculated because teamwork competencies were rated using a continuous scale. The Kappa coefficient corrects for agreement that was due to chance, but is only appropriate for categorical data (Pallant, 2007). In general, the percent of overall agreement was substandard for the ratings of teamwork competencies, ranging from 44% to 80% across each task and for both tasks combined. *Contingency Planning* consistently demonstrated the worst agreement. The percentage of scores within a single point, however, was nearly perfect, ranging from 80% to 100% across rounds and handoffs, handoffs, and rounds. This means raters interpreted team competencies in a similar way, but ultimately scored behaviors differently.

Given the findings reported about the usage of the scale in the previous section, a logical consideration is whether the process of measurement actually involved a three point scale or a five point scale (i.e., considering values from the midpoint to the positive endpoint vs. the entire scale). If a three point scale is conceptualized, the unfavorable agreement is magnified even further and the closeness of scoring attributions between raters is inconsequential. It is important to reiterate that the negative end of the scale was used, albeit scarcely. This finding provides confidence that raters were indeed considering the full range of the scale when making judgments of team behaviors. The issue of range restriction, however, increases the concern over attribution disagreements. A key consideration for improving rater agreement is to provide resources to help raters better discriminate the quality of team behaviors, especially at the intermediary points of the scale. This can be accomplished through rater training and developing guidance materials that provide explicit examples of why certain behaviors should be scored in a particular way for particular tasks. Another area to improve rater agreement could be to revise the scale itself to include behavioral anchors at the intermediary points in addition to the end points and midpoints to guide measurement.

Future Development and Testing

This study developed and tested the psychometric properties of a behavioral marker system rooted in theory and relevant to critical care. A main consideration is that only a subset of teamwork competencies were assessed for a subset of tasks. Future efforts will be dedicated to evaluating the validity of the marker system during action-oriented cycles of team task accomplishment that emphasize different teamwork competencies. Studies are planned to observe mobility sessions and admissions in the same unit with the same methodological approach to this end (i.e., two raters making judgments of relevant teamwork behaviors).

Similarly, the full structure of the marker system should be tested. The current study only required raters to score subdimensions of a subset of teamwork behaviors. Future studies should establish the best scoring technique at the dimension level as well (e.g., weighting of subdimensions for particular tasks, different scale points). This consideration is especially important for observations of successive instances of teamwork (e.g., rounds and when a nurse is assigned two patients for handoffs). It may prove too difficult for raters to take the extra time to make judgments of behavior at multiple levels given the paucity of time between instances of teamwork. This may leave the rater scoring a previous instance of teamwork as another instance begins. In such cases, the raters should not rate the subsequent instance of teamwork because they are not attending to relevant performance. Scoring attributions at the dimension level would also provide another facet to explore in a G study as described earlier, with subdimensions being nested within dimensions.

As more data are collected, different statistical approaches for psychometric analysis can also be leveraged. Notably, data collected in future validation efforts will be subjected to exploratory factor analysis (EFA), which is a large sample size technique. EFA was applied in the psychometric evaluation of the NOTSS marker system described in Chapter 3 (Crossley et al., 2011). The authors tested the internal structure of the tool by examining how well the elements (i.e., subdimensions) of the rating tool conformed to the hierarchical structure of the instrument (i.e., domains). A unique contribution planned in future psychometric testing of the ICU behavioral marker system will be contrasting the internal structure of the marker system when data from rounds, handoffs, and other action-oriented team tasks are combined compared to when they are looked at separately. Such an analysis would reveal whether the structure of the marker system is consistent across all team tasks, or whether subdimensions load differently depending on what task is examined.

This study provided initial evidence that the marker system can have utility in differentiating among teamwork competencies. Before the tool can be leveraged in research and practice, however, mechanisms for improving rater reliability and agreement are needed. First, the rating scale may need to be revised. Absolute agreement between raters was not favorable, though raters generally scored behaviors within a single point of each other and did not systematically make different scoring attributions. This may suggest the number of scale points should be reduced, especially taking into account that the negative end of the scale was scarcely used by either rater. Reducing the number of scale points from five to four may increase rater reliability without greatly comprising the sensitivity of ratings. Before adjusting the scale, however, observations of teamwork in other units without structured protocols is warranted. Additionally, future studies of action oriented tasks may demonstrate different scale usage.

Another structural change to the rating scale could be the addition of a frequency count to keep track of instances of good and poor teamwork to supplement the notes column. During a single observation, there may be several examples of both good and poor manifestations of the same teamwork competency. Retaining information related to fluctuations in performance is a key challenge for raters when making global assessments (Gaba et al., 1998). That said, frequency counts should not singularly be used to make the final decision for a rating; some occurrences of the behavior may be more meaningful for scoring attributions than others and rater judgment is needed to make that determination.

Aside from structural changes to the marker system, more explicit rater guidance and training materials should be developed to improve rater reliability and agreement. A key consideration is to develop low-cost resources to supplement instructor-based training. Scenario vignettes will be developed along with a scoring guide that explains the rationale for why behaviors should be rated a certain way to this end. Particular emphasis should be given to areas where rater reliability was less than ideal, as described earlier in this discussion Another strategy could be to train a group of 'super users' in the tool. These ICU behavioral marker system experts could provide additional guidance through activities such as coaching calls or web conferences to supplement in person training, though it is worth acknowledging this is a consideration for when the tool is more established and ready to be used for research or applied purposes.

A final area that merits future consideration is to examine whether there are systematic differences in how teamwork is scored based on rater background. Clinicians should be reengaged and trained in the use of the marker system. A generalizability study could detect whether there are systematic differences in how subdimensions are scored not just by rater background (e.g., clinician vs. non-clinician) but also by clinician type (e.g., nurse vs. physician). Systematic variance would indicate that certain types of raters score behaviors differently. Ultimately, users from a diverse array of backgrounds (e.g., human factors, clinical) should be able to reliably use the marker system. Unwanted variance may suggest areas for future system development so that a wider range of raters with different professional backgrounds can adequately use the tool.

Conclusion

There is a burgeoning body of work dedicated to understanding teamwork in the ICU. Behavioral marker systems offer an objective strategy for the assessment of teamwork, yet no system existed to quantify focal teamwork skills across a wide range of ICU tasks prior to the present study. Team performance measurement has a number of practical applications such as structuring feedback to guide learning and assessing teamwork skills on the job; therefore, future efforts to improve teamwork in this setting (e.g., through training) can leverage the marker system to enumerate whether the goals and objectives of the intervention realized (Rosen et al., 2010). The marker system is also linked to prominent teamwork theory, ensuring that nomenclature used to describe teamwork competencies is accurate. A psychometric approach to validity testing was also applied, affording the examination of areas to improve the use of the tool in future applications. Subsequent studies exploring the validity of the tool in a broader range of team tasks and for a wider variety of teamwork competencies are planned.

APPENDIX A: STUDY BACKGROUND (TEAMWORK IN ICU REVIEW)

Citation	**Type of Article** 1. Empirical-Quantitative 2. Empirical-Qualitative	**Teamwork Process Investigated** (e.g., from methods/results)	**Team Task Investigated** (e.g., from methods/results)	**Teamwork Intervention Investigated/Observed**	**Was the Intervention Developed Primarily (ONLY) for Teamwork?** Y/N/NA	**Type of Experiment** 1. True Experiment 2. Quasi-Experiment 3. Non-Experimental	**Description of Organization**
Meurling et al. (2013)	1	-Communication -Collaboration -Team Climate	Clinical Tasks: (Five standardized scenarios were pre-programmed: urosepsis, pneumothorax, aortic rupture with re-bleeding after operation, anaphylaxis due to administration of a drug in the ICU and hypovolaemia due to ruptured spleen).	Combination lecture and SBTT; high-fidelity, structured team coordination training; The strategy for collaboration used during SBTT, the all team members' behavior (A-TEAM) programmed	Y	2	General ICU in a Swedish University hospital (both pediatric and adult patients)
Miller and Buerhaus (2013)	1	Team Coordination (preparation, planning, direction, execution, and team assessment)	-Handoff	Not Specified	N/A	3	The hospital has eight specialist ICUs: burn, cardiovascular, medical, neurological, neonatal, pediatric, surgical, and trauma ICUs, with between 40 and 60 beds each.
Douglas et al. (2013)	1	Care Coordination	Handoff (shift change) Rounds	Not Specified	N/A	3	A medical-surgical AICU, a cardiac ICU (CICU), a PICU, and a neonatal ICU (NICU) at a 400-bed tertiary care community teaching hospital.

Citation	**Type of Article** 1. Empirical-Quantitative 2. Empirical-Qualitative	**Teamwork Process Investigated** (e.g., from methods/results)	**Team Task Investigated** (e.g., from methods/results)	**Teamwork Intervention Investigated/Observed**	**Was the Intervention Developed Primarily (ONLY) for Teamwork?** Y/N/NA	**Type of Experiment** 1. True Experiment 2. Quasi-Experiment 3. Non-Experimental	**Description of Organization**
Figueroa et al. (2012)	1	-Closed loop communication -Mutual respect -Empowerment	-Clinical; Post pediatric cardiac surgery cardiac arrest management -Huddles/Debriefs	-SBTT, TeamSTEPPS -incorporated lecture and high fidelity sim followed by structured debrief	Y	2	Pediatric cardiac ICU (PCICU); Study was conducted at an off-site sim center
Newkirk et al. (2012)	1	Communication	Rounds	-Checklist	Y	2	-Academic military center -20 bed surgical trauma ICU (STICU) -16-bed burn ICU (BICU)
Rincon et al. (2012)	1	Team Satisfaction	rounds	-Robotic Tele-Presence (RTP)	Y	2	-26-bed Neuro-ICU
Jukkala et al. (2012)	1	-Communication openness -Communication quality	Handoff	-Shift report (MICU Communication Tool)	Y	2	-Academic health center -25-bed medical ICU
Karanikol et al. (2012)	1	Collaboration (article discusses perceived quality of professional interaction in methods)	Not Specified	Not Specified	N/A	3	-ICUs in general hospitals in public and private sectors in Greece
Vigorita et al. (2011)	1	Teamwork climate	Not Specified	Safety Attitude Questionnaire Action Plan	N	2	All ICUs in Rhode Island

Citation	**Type of Article** 1. Empirical-Quantitative 2. Empirical-Qualitative	**Teamwork Process Investigated** (e.g., from methods/results)	**Team Task Investigated** (e.g., from methods/results)	**Teamwork Intervention Investigated/Observed**	**Was the Intervention Developed Primarily (ONLY) for Teamwork?** Y/N/NA	**Type of Experiment** 1. True Experiment 2. Quasi-Experiment 3. Non-Experimental	**Description of Organization**
Stocker et al. (2012)	1	Teamwork Communication	Clinical task; Pediatric resuscitation	SPRinT, simulated pediatric resuscitation team training; High fidelity; Included debrief -Debrief focused on teaching CRM principles	Y	3	paediatric intensive care unit (PICU)
Stone et al. (2011)	1	Not Specified	Rounds	-Daily Goal Rounds -Checklist	NS	2	12-bed surgical intensive care unit (SICU)
Mayer et al. (2011)	1	-Communication -leadership -situation monitoring mutual support/assertion overall teamwork overall leadership	Clinical; Extracorporeal membrane oxygenation (ECMO)	TeamSTEPPS, classroom-based	Y	2	Surgical ICU, 16-bed Pediatric ICU, 20-bed

Citation	Type of Article 1. Empirical-Quantitative 2. Empirical-Qualitative	Teamwork Process Investigated (e.g., from methods/results)	Team Task Investigated (e.g., from methods/results)	Teamwork Intervention Investigated/Observed	Was the Intervention Developed Primarily (ONLY) for Teamwork? Y/N/NA	Type of Experiment 1. True Experiment 2. Quasi-Experiment 3. Non-Experimental	Description of Organization
Pascual et al. (2011)	1	Teamwork (Communication Task delegation Leadership); Situation awareness (calmness/assertiveness, team/distraction management) *there were others, but not necessarily directly relevant to teamwork	Clinical, (1) anaphylaxis with tension pneumothorax, (2) septic shock from Clostridium difficult colitis, (3) myocardial infarction [MI] with diabetic ketoacidosis, (4) hemorrhagic shock with abdominal compartment syndrome, and (5) deteriorating traumatic brain injury with status epilecticus.	SBT, high fidelity	Y	2	Surgical ICU
Chang et al. (2010)	1	Communication (openness and effectiveness) Leadership	NS	Not Specified	N/A	3	ICU
Reader et al (2011)	1	communication information sharing Team SA collaboration Leadership (communication, team SA, leadership style)	Rounds	Not Specified	N/A	3	16-bed ICU

Citation	**Type of Article** 1. Empirical-Quantitative 2. Empirical-Qualitative	**Teamwork Process Investigated** (e.g., from methods/results)	**Team Task Investigated** (e.g., from methods/results)	**Teamwork Intervention Investigated/Observed**	**Was the Intervention Developed Primarily (ONLY) for Teamwork?** Y/N/NA	**Type of Experiment** 1. True Experiment 2. Quasi-Experiment 3. Non-Experimental	**Description of Organization**
Rehder et al. (2012)	1	-Communication barriers Team culture Shared goal agreement	Daily rounds	1) a new resident daily progress note format; 2) performance improvement “dashboard” 3) use of a bedside whiteboard to document daily goals	Y	2	16-bed pediatric intensive care unit PICU
Samuels et al. (2011)	1	Not Specified	Clinical; Treating Patients with Aneurysmal Subarachnoid Hemorrhage	Specialized staffing (Dedicated Neurocritical Care Team)	N	2	neuroscience intensive care units (NICU)
Sexton et al. (2011)	1	Teamwork climate (subsection of SAQ)	Not Specified	CUSP	N	2	71 ICUs participating in the Michigan Health and Hospital Association Keystone ICU Project
Ahmed et al. (2012)	1	communication	Clinical rounds	Specialized staffing (Dedicated Central Catheter Team) Rounds (Daily Discussion of CVC Necessity)	N	2	pediatric intensive care unit (PICU) PICU, 30-bed
Palma et al. (2011)	1	Not Specified	Handoff	-Integration of handoff tool with EMR	N	2	74-bed Newborn ICU) NICU
Collins et al. (2010)	1	Not Specified	rounds	Not Specified	N/A	3	neurovascular ICU (NICU)

Citation	**Type of Article** 1. Empirical-Quantitative 2. Empirical-Qualitative	**Teamwork Process Investigated** (e.g., from methods/results)	**Team Task Investigated** (e.g., from methods/results)	**Teamwork Intervention Investigated/Observed**	**Was the Intervention Developed Primarily (ONLY) for Teamwork?** Y/N/NA	**Type of Experiment** 1. True Experiment 2. Quasi-Experiment 3. Non-Experimental	**Description of Organization**
Vats et al. (2010)	1	communication	rounds	Not Specified	N/A	3	pediatric intensive care unit. During the time of analysis, The new unit increased from 21 beds to 30 beds and grew from 11,000 square feet to 33,000 square feet.
Adler-Milstein et al. (2011)	1	Communication (openness, accuracy, timeliness, and satisfaction Coordination Membership stability Collaborative decision-making	Not Specified	Not Specified	N/A	3	6 adult ICUs
Allan et al., (2010)	1	Confidence in leading future code events Speaking up	Clinical Tasks; Sims were based on actual cases	CRM, classroom-based, high-fidelity sim, debriefing; discovery based gameplay prior to classroom lesson to learn about CRM skills	Y	3	24-bed pediatric cardiac intensive care unit (pCICU)
LeBlanc et al. (2010)	1	Team commitment Collaborative practice (open communicating, cooperative problem solving, professional recognition)	Not Specified	Not Specified	N/A	3	ICUs from 8 different European countries

Citation	Type of Article 1. Empirical-Quantitative 2. Empirical-Qualitative	Teamwork Process Investigated (e.g., from methods/results)	Team Task Investigated (e.g., from methods/results)	Teamwork Intervention Investigated/Observed	Was the Intervention Developed Primarily (ONLY) for Teamwork? Y/N/NA	Type of Experiment 1. True Experiment 2. Quasi-Experiment 3. Non-Experimental	Description of Organization
Kim et al. (2010)	1	Not Specified	Rounds, but unable to investigate	Not Specified	N/A	3	112 hospitals and 107 324 patients in the final analysis
Johnson et al. (2009)	1	Not Specified	Rounds	Multidisciplinary Rounds	NS	2	ICU
Pronovost et al. (2008)	1	Teamwork Climate (scale items in table 4, p. 214)	Not Specified	CUSP	N	2	72 ICUs
Stockwell et al. (2007)	1	Leadership/management (physicians)	rounds	Not Specified	N/A	3	pediatric ICU (PICU); 24-bed
Phipps and Thomas (2007)	1	communication	rounds	Daily goal sheet	Y	2	Medical-surgical pediatric ICU; 12-bed
Pronovost et al. (2006)	1	Communication team structure Leadership	-Handoff -Clinical, routine care	Not Specified	N/A	3	23 ICUs were actively reporting at the time of this analysis
Huang et al. (2007)	1	Teamwork climate (perceived quality of collaboration)	Not Specified	Not Specified	N/A	3	4 ICUs with a range from 10-30 beds

Citation	**Type of Article** 1. Empirical-Quantitative 2. Empirical-Qualitative	**Teamwork Process Investigated** (e.g., from methods/results)	**Team Task Investigated** (e.g., from methods/results)	**Teamwork Intervention Investigated/Observed**	**Was the Intervention Developed Primarily (ONLY) for Teamwork?** Y/N/NA	**Type of Experiment** 1. True Experiment 2. Quasi-Experiment 3. Non-Experimental	**Description of Organization**
Jain et al. (2006)	1	Not Specified	Not Specified	(1) physician led multidisciplinary rounds; (2) daily ''flow'' meeting to assess bed availability; (3) ''bundles'' (sets of evidence based best practices); (4)culture changes with a focus on the team decision making process.	N (ventilator and central line bundle)	2	28 bed Medical-Surgical ICU
Narasimhan et al. (2006)	1	communication	Rounds	Daily goal sheet	Y	2	16-bed ICU
Sluiter et al. (2005)	1	communication	Not Specified	Multidisciplinary work shift evaluations	Y	2	pediatric intensive care unit (PICU) in The Netherlands
Boyle and Kochinda (2004)	1	Collaborative Communication Leadership Communication Coordination Problem solving & conflict management Team culture Cohesion	Not Specified	Collaborative Communication Intervention (contains modules, but not sure if lecture-based or sim-based or mixed)	Y	2	1 ICUs from 2 different hostpitals

Citation	Type of Article 1. Empirical-Quantitative 2. Empirical-Qualitative	Teamwork Process Investigated (e.g., from methods/results)	Team Task Investigated (e.g., from methods/results)	Teamwork Intervention Investigated/Observed	Was the Intervention Developed Primarily (ONLY) for Teamwork? Y/N/NA	Type of Experiment 1. True Experiment 2. Quasi-Experiment 3. Non-Experimental	Description of Organization
Wheelan et al. (2003)	1	Not Specified (stages of group development)	Not Specified	Not Specified	N/A	3	17 ICUs in 9 hospitals
Dodek and Raboud (2003)	1	Communication	Rounds	Explicit Approach to Rounds	Y	2	15-bed ICU
Studdert et al. (2003)	1	Communication Leadership Coordination involvement in decisions	Life-sustaining treatment	Not Specified	N/A	3	7 ICUs
Pronovost et al. (2003)	1	Communication	Rounds	Daily Goals Form	Y	2	16-bed surgical ICU
Pollack and Koch (2003)	1	-Teamwork and leadership '-Relationships and communication (openness, accuracy, timeliness) -Coordination conflict resolution	Not Specified	Not Specified	N/A	3	Eight acute care neonatal intensive care units in Washington, DC
Thomas et al. (2003)	1	Team climate (ability to speak up, collaboration, cooperation, coordination, support, conflict resolution, decision-making input)	Not Specified	Not Specified	N/A	3	Eight nonsurgical intensive care units

Citation	**Type of Article** 1. Empirical-Quantitative 2. Empirical-Qualitative	**Teamwork Process Investigated** (e.g., from methods/results)	**Team Task Investigated** (e.g., from methods/results)	**Teamwork Intervention Investigated/Observed**	**Was the Intervention Developed Primarily (ONLY) for Teamwork?** Y/N/NA	**Type of Experiment** 1. True Experiment 2. Quasi-Experiment 3. Non-Experimental	**Description of Organization**
Miller (2001)	1	Items used to gauge collaborative interaction: Leadership communication openness communication satisfaction communication timeliness problem solving	Multidisciplinary Meetings [very small component of research scope]	Not Specified	N/A	3	22-bed medical-surgical ICU
Dimick et al. (2001)	1	Not Specified	Rounds with an ICU physician Clinical; esophageal resection	Not Specified	N/A	3	Nonfederal acute care hospitals in Maryland that performed esophageal resection (n=35 hospitals) during the study period, 1994–1998
Henneman et al. (2001)	1	Collaboration Communication	Clinical, Weaning from mechanical ventilator Rounds (discussion of weaning plan)	Collaborative weaning plan including a weaning board and flow sheet	Y	2	8-bed medical ICU
Baggs et al. (1999)	1	Collaboration	Transfer (intra unit decision to transfer out of unit)	Not Specified	N/A	3	1 Medical ICU (20-bed), 1 Surgical ICU (16-bed), 1 mixed ICU (7-bed)

Citation	**Type of Article** 1. Empirical-Quantitative 2. Empirical-Qualitative	**Teamwork Process Investigated** (e.g., from methods/results)	**Team Task Investigated** (e.g., from methods/results)	**Teamwork Intervention Investigated/Observed**	**Was the Intervention Developed Primarily (ONLY) for Teamwork?** Y/N/NA	**Type of Experiment** 1. True Experiment 2. Quasi-Experiment 3. Non-Experimental	**Description of Organization**
Donchin et al. (1995)	1	Communication (verbal)	Rounds Handoff/Shift change (activities are not necessarily associated with these outcomes, but these were mentioned in methods)	Not Specified	N/A	3	Medical -surgical ICU (Jerusalem)
Shortell et al. (1994)	1	Caregiver interaction: '-Culture -leadership -communication -coordination -problem-solving/conflict management	Not Specified	Not Specified	N/A	3	42 medical and surgical ICUs
Baggs and Ryan (1990)	1	Collaboration	Transfer (intra unit decision to transfer out of unit)	Not Specified	N/A	3	Medical ICU
Ilan et al. (2012)	1	Communication	Handoff	Not Specified	N/A	3	-The ICU is an academic 21-bed unit providing care for medical, surgical, trauma and cardiovascular surgery patients

Citation	**Type of Article** 1. Empirical-Quantitative 2. Empirical-Qualitative	**Teamwork Process Investigated** (e.g., from methods/results)	**Team Task Investigated** (e.g., from methods/results)	**Teamwork Intervention Investigated/Observed**	**Was the Intervention Developed Primarily (ONLY) for Teamwork?** Y/N/NA	**Type of Experiment** 1. True Experiment 2. Quasi-Experiment 3. Non-Experimental	**Description of Organization**
Abraham et al (2012)	2	Communication Information Exchange Coordination Collaborative problem-solving	Handoff	Not Specified	N/A	3	16-bed medical ICU
Alvarez and Coiera (2005)	2	communication	Rounds	Not Specified	N/A	3	12-bed ICU in Australia
Baggs and Schmitt (1997)	2	Collaboration Communication	Not Specified	Not Specified	N/A	3	Medical intensive care unit
Collins et al (2012)	2	Coordination Communication	Handoff	Not Specified	N/A	3	-21 bed Cardiothoracic Intensive Care Unit (CTICU)
Collins et al. (2011)	2	information exchange communication shared mental model Collaborative decision-making coordination	rounds handoffs	Not Specified	N/A	3	18-bed neurovascular ICU (NICU)
Custer et al. (2012)	2	Communication Shared mental model	Rounds Handoff	Not Specified	N/A	3	26-bed pediatric intensive care unit (PICU)

Citation	Type of Article 1. Empirical-Quantitative 2. Empirical-Qualitative	Teamwork Process Investigated (e.g., from methods/results)	Team Task Investigated (e.g., from methods/results)	Teamwork Intervention Investigated/Observed	Was the Intervention Developed Primarily (ONLY) for Teamwork? Y/N/NA	Type of Experiment 1. True Experiment 2. Quasi-Experiment 3. Non-Experimental	Description of Organization
Hawryluck et al. (2002)	2	Collaboration Communication Leadership/authority Cohesion	-Rounds -Clinical (managing a feeding tube)	Not Specified	N/A	3	2 ICUs
Jirapaet et al. (2006)	2	communication	handoff	Not Specified	N/A	3	4 large neonatal intensive care units in Thailand
Lingard et al. (2004)	2	Collaboration	round	Not Specified	N/A	3	ICU team members in two urban teaching hospitals in Toronto, Canada
Ohlinger et al. (2003)	2	-Communication -Collaboration -Conflict management -Coordination -Leadership -Shared problem solving	Not Specified	Not Specified	N/A	3	4 neonatal intensive care units (NICUs).
Reader et al. (2011)	2	Team leadership Functional leadership behaviors: listed in table 3, p. 1685)	-Rounds	Not Specified	N/A	3	7 General ICUs
Rosengren et al. (2007)	2	Leadership (nursing)	Not Specified	Not Specified	N/A	3	ICU in Sweden with 10 beds
Heffner et al. (1996)	1	Communication	Resuscitation	Structured, procedure specific DNR order forms	Y	2	Neurosurgical ICU

Citation	**Type of Article** 1. Empirical-Quantitative 2. Empirical-Qualitative	**Teamwork Process Investigated** (e.g., from methods/results)	**Team Task Investigated** (e.g., from methods/results)	**Teamwork Intervention Investigated/Observed**	**Was the Intervention Developed Primarily (ONLY) for Teamwork?** Y/N/NA	**Type of Experiment** 1. True Experiment 2. Quasi-Experiment 3. Non-Experimental	**Description of Organization**
King and Lee (1994)	1	Collaboration	Not Specified	Not Specified	N/A	3	ICUs at the Navy's 4 teaching hospitals and 2 hospital ships
Wayne et al. (2008)	1	Communication	Handoff	Standardized Patient Handoff System (spreadsheet)	N	2	-5 Cardiovascular ICUs -Surgical ICU -3 surgical floors (single hospital)
Stockwell et al. (2005)	1	Leadership	Not Specified	Not Specified	N/A	3	N/A (pediatric critical care review board course)
Weller et al. (2011)	1	-Leadership and Team Coordination -Mutual performance monitoring -Verbalizing situational information	Clinical; sim scenarios involved airway and cardiovascular emergencies	Training on crisis resource management	Y	2	9 CCUs
Sneve et al. (2008)	1	Not Specified	Rounds	Staffing; Multidisciplinary team establishment (that includes a registered dietician)	N	2	Neonatal ICU
DuBose et al. (2010)	1	Not Specified	Rounds	Quality rounds checklist	N	2	Trauma ICU
Wright et al. (1996)	1	Communication	Rounds	Nurse presentation with nursing prompt sheet	N	2	Royal Hallamshire
Boos et al. (2010)	2	Communication Cohesion	Clinical; Tracheostomy	Comprehensive Care Rounds	Y	3	Neonatal Intensive Care Unit

Citation	Type of Article 1. Empirical-Quantitative 2. Empirical-Qualitative	Teamwork Process Investigated (e.g., from methods/results)	Team Task Investigated (e.g., from methods/results)	Teamwork Intervention Investigated/Observed	Was the Intervention Developed Primarily (ONLY) for Teamwork? Y/N/NA	Type of Experiment 1. True Experiment 2. Quasi-Experiment 3. Non-Experimental	Description of Organization
Rangachari et al. (2010)	2	Communication	Not Specified	Not Specified	N/A	3	4 ICUs (medical, surgical, neonatal, and pediatric), single hospital
Manias and Street (2001)	2	Communication	Rounds	Not Specified	N/A	3	Single Critical Care Unit
Manias and Street (2000)	2	Communication	Handoff	Not Specified	N/A	3	Single Critical Care Unit
Collins et al. (2010)	2	Communication	Rounds	Not Specified	N/A	3	Neurovascular ICU
Ho et al. (2007)	2	Communication Information Exchange Joint Sense-Making	Rounds	Not Specified	N/A	3	Pediatric ICU
Pickering et al. (2009)	1	Not Specified	Handoff	Standardized presentation format	N	2	Medical-Surgical ICU
Miller et al. (2009)	1	Coordination	Handoff Round	Not Specified	N/A	3	General ICU
Miller et al. (2009)	1	Communication	Handoff Round	Not Specified	N/A	3	2 ICUs
O'Connor et al. (2009)	1	Communication	Not Specified	Wireless email	Y	3	Medical-Surgical ICU
Nunnink et a. (2009)	1	Not Specified	Clinical; emergency chest reopen	Simulation-based Team Training; high fidelity	Y	2	Surgical ICU
Cardarelli et al. (2009)	2	Not Specified	Rounds	Not Specified	N/A	3	Pediatric ICU

Citation	**Type of Article** 1. Empirical-Quantitative 2. Empirical-Qualitative	**Teamwork Process Investigated** (e.g., from methods/results)	**Team Task Investigated** (e.g., from methods/results)	**Teamwork Intervention Investigated/Observed**	**Was the Intervention Developed Primarily (ONLY) for Teamwork?** Y/N/NA	**Type of Experiment** 1. True Experiment 2. Quasi-Experiment 3. Non-Experimental	**Description of Organization**
Vivian et al. (2009)	2	Trust Cooperation	Not Specified	Not Specified	N/A	3	Pediatric ICU
Piquette et al. (2009)	2	Coordination Collaboration Leadership Conflict Management	Not Specified	Not Specified	N/A	3	4 ICUs at a single hospital
Linton et al. (2009)	2	Leadership	Not Specified	Not Specified	N/A	3	General ICU

APPENDIX B: STUDY RESULTS (TEAMWORK IN ICU REVIEW)

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
Meurling et al. (2013)	1. Improved perception of team climate among nurse assistants 2. Improved perception of collaboration and communication with physicians among nurse assistants	NS	1. Improved self-efficacy among physicians and nurses	1. Reduced nurse turnover 2. Reduction in sick leave among nurses and nurse assistants	1. Increased perceived quality of safety climate among nurses and nurse assistants	N
Miller and Buerhaus (2013)	NS	NS	1. The type of support tool used depends on the phase of team coordination charge nurses were engaged in. 2. The patient list was used by CNs during handoffs and updates with other staff nurses 3. 29% of coded units involved CN nurse interaction with other peers 4. 22% of coded units involved CN nurse interaction with other team members 5. CN generally do not rely on support tools during execution and assessment decisions involving staff and other team members	NS	NS	Y

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
Douglas et al. (2013)	NS	NS	1. Nurses were observed to spend 52% of time in direct patient care, 23% care coordination, 2% on direct patient care activities, and 24% on non patient care 2. Nurses only spent 3% of time talking with physicians 3. Nurses were engaged in coordination tasks 79% of the time during handoffs (21% outside handoffs) 4. Nurses were engaged in coordination tasks 80% of the time during rounds (22% outside rounds) 5. On average, nurses switched between tasks every 29 seconds (125 activities per hour).	NS	NS	N
Figueroa et al. (2012)	-Increased confidence and skill in leading code events -Increased use of closed-loop communication -Increased perception of mutual respect and empowerment	NS	-Increase confidence and skill in advanced airway management and cardioversion/defibrillation -Increased use of huddles/debriefs	NS	NS	N
Newkirk et al. (2012)	NS	NS	-Checklist items were discussed more frequently after the checklist was implemented	NS	NS	N
Rincon et al. (2012)	-Increased team satisfaction among	NS	NS	NS	NS	Y

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
	nurses					
Jukkala et al. (2012)	-Improved overall perception of communication handoff communication (quality and openness of communication did not show a significant improvement)	NS	NS	NS	NS	N
Karanikol et al. (2012)	NS	NS	NS	1. Approx. 21% of participants reported moderate anxiety symptoms, with sleep disturbance being mentioned the most 2. Satisfaction of professional interaction between nursing personnel and phycisians was negatively associated with anxiety, anxious mood, tension, depression, muscular, cardiovascular, and genitourinary symptoms, and sleep disturbances 3. Satisfaction of professional interaction among nurses was negatively associated with anxiety, anxious	NS	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
				mood, tension, sleep disturbances, depression, and cardiovascular symptoms 4. Satisfaction of professional interaction did not significantly predict higher anxiety		
Vigorita et al. (2011)	1. Units with SAQAP observed a greater increase in improved team climate scores, but this difference was not statistically significant.	1. Units with SAQAP decreased CLABSI rates by approx. 10% compared to approx. 2% in units without SAQAP, but this difference was not significant 2. Units with SAQAP decreased VAP rates by approx. 15% compared to approx. 5% in units without SAQAP, but this difference was not significant	NS	1. Units with SAQAP observed greater improvement in job satisfaction and stress recognition, but these differences were not statistically significant	1. Units with SAQAP observed greater improvement in safety climate and perceptions of management, but these differences were not statistically significant	Y

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
Stocker et al. (2012)	1. Impact of SPRinT on teamwork and communication was perceived as effective by approx. 91% of respondents 2. The perceived effectiveness of SPRinT on non-technical skills was significantly sustained from the introduction to intermediate phase and the intermediate phase to the established phase.	NS	1. Impact of SPRinT on technical skills was perceived as effective among approx. 70% of respondents 2. The perceived effectiveness of SPRinT on technical skills was significantly sustained from the introduction to intermediate phase and the intermediate phase to the established phase.	NS	NS	N
Stone et al. (2011)	NS	1. The number of VAPs significantly decreased by 67% following the implementation of GR 2. There was a significant decrease in the incident rate from 26.8 VAPs to 7.0 VAPs per 1,000 ventilator days. 3. There was not a significant difference in patient mortality 4. There was not a significant different in mean ventilator days 5. There were decreases in the average SLOS and HLOS, but these	NS	NS	NS	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
		differences were not significant				
Mayer et al. (2011)	1. 1 month following implementation, there were significant improvements in communication, leadership, situation monitoring, mutual support, overall teamwork, and overall leadership 2. Communication, leadership, situation monitoring, and overall teamwork did not remain significantly improved 6 months after the intervention 3. Small changes were noticed in the nurses perception of teamwork, both for	1. The rate of nosocomial infection decreased for all but 4 months	1. The amount of time between the decision to place a patient on EMCO and the placement of the patient on EMCO was significantly lower.	NS	NS	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
	RN-RN interactions and RN-MD interactions					
Pascual et al. (2011)	1. Improvements were observed among all trainees for teamwork, decision making, and SA 2. Leadership skills were significantly greater for fellows than AP's before training, but became similar following training	NS	1. Improvements were observed in emergency clinical skills post intervention, but not significantly 2. Multiple choice performance improved significantly for advanced practitioners, but not fellows	NS	NS	Y

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
Chang et al. (2010)	1. Communication was perceived to be significantly more open by doctors (73%) than nurses (32%) 2. Doctors perceived doctor-doctor comms and doctor to nurse comms as good. Nurses perceived there was less comms between doctors and nurses 3. 61% of nurses and 50% of Drs. reported Dr-Nurse communication as effective across shifts.	NS	1. 53% Drs reported other physicians had given them incorrect patient information. 67% felt nurses had given them incorrect patient information, 2. 32% of nurses reported that Drs had given them incorrect patient information. 51% felt nurses felt other nurses had given them incorrect patient information 3. 20% of Drs. And 39% of nurses considered it necessary to recheck information they had received. 4. 88% of nurses felt they called Drs in a timely manner regarding patient care. 53% of Drs. felt nurses called them in a timely manner regarding patient care. 5. Compared to Dr. (63%), nurses (87%) felt they had an overall understanding of patient care goals. 6. 28% of Drs felt the ICU always meets patient care treatment goals, compared to 65% of nurses 7. Nurses tended to think treatment outcomes were	1. Nurses did not consider it enjoyable to talk with Drs. 2. Nurses did not consider it enjoyable to talk with other nurses, and specifically sisters 3. Drs. Considered it enjoyable to talk with each as well as nurses 4. Drs reported they felt it was easy to receive advice from other Drs. (90-100%) 5. More Drs (83%) than nurses (63%) reported they felt that it was easy to take advice from senior nurses	1. Only 3% of physicians feel the most sophisticated technology is applied to patient care as compared to 53% of nurses	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
			more favorable (73%) than what Drs felt (48%)			
Reader et al (2011)	1. Verbal communications made by nursing staff, senior trainees, and junior trainees correlated with self-ratings of involvement 2. Verbal contributions by nurses, senior trainees, and junior trainees were positively associated with Snr. Dr. prompts 3. Snr Dr./Snr. trainee team SA of patient deterioration was	NS	1. Sr. Drs. were significantly more accurate anticipating patient deterioration likelihood than Jr. Drs. 2. Anticipations were correct 71% of the time when all team members formed shared SA 3. Sr. Dr. anticipations were incorrect 58% of the time when they were in a minority	NS	NS	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
	predicted by Snr. trainee involvement during decision-making 4. Snr. Dr./Jnr trainee team SA for patient deterioration was predicted by Jnr. trainee involvement					
Rehder et al. (2012)	-Mean team agreement improved from approx. 57% before the intervention to approx. 83% after the intervention 2. The number of clinicians who considered themselves as a valuable member of the team increased from approx. 59% before the intervention to approx. 77% after the intervention.	NS	1. Barriers to communication were reduced after the intervention 1A. Nurse bedside multitasking (approx. 28% to approx. 9%) 1B. Interruptions during patient presentations (approx. 17% to 8%) 1C Group disassociation (approx.. 17% to approx. 7%) 2. Use of facilitators to communication were increased 2A. Review of prior daily goals (approx. 1% to approx. 94% 2B. Solicitation of bedside nurse input (approx. 74% to 97%) 2C. Confirmation of patient orders at conclusion of patient discussion (approx.	1. Approx. 43% of physicians were 'very satisfied' or 'satisfied' with the rounding process prior to the intervention, compared to approx. 78% after the intervention	NS	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
			76% to approx. 93%) 3. No differences were observed in the time required to complete rounds			
Samuels et al. (2011)	NS	1. Patients were more likely to be discharged to their home following the intervention (approx. 37%) than before (approx. 25%). 2. Patients were less likely to be discharged to a rehab facility following the intervention (approx. 32%) than before the intervention (approx. 43%) 3. Patients were more likely to receive definitive aneurysm treatment following the intervention (18%) than before the intervention (approx. 11%).	NS	NS	NS	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
Sexton et al. (2011)	NS	NS	NS	NS	1. Improved safety climate scores were observed following the intervention (approx. 43% to approx. 52%) 2. Less ICUs needed significant safety climate improvements after the intervention (33) than before (62) 3. Safety climate improved significantly across a variety of hospital characteristics (e.g., bed size, teaching, nonteaching, faith-based, not faith-based)	N
Ahmed et al. (2012)	NS	1. CA-BSI rate was reduced from 7.9 infections per 1000 central catheter days to 1.3 infections per 1000 central catheter days (reduction of approx..	NS	NS	NS	No Limitations Specified

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
		84%)				
Palma et al. (2011)	NS	NS	1. The perceived accuracy of the sign-out document increased following the implementation of the intervention (78% to 91%) 2. The satisfaction with the process of updating information in the sign-out document increased following the intervention (35% to92%) 3. The satisfaction with the printed sign-out document improved following the intervention (71% to 98%) 4. More time was spent updating sign-out information after the intervention (16-20min) than before (11-15min) 5. Time spent transcribing EMR data during sign-out preparation reduced from 25-49% before the intervention to less than 25% following the intervention.	NS	NS	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
Collins et al. (2010)	NS	NS	1. Approx. 24 % of goals stated during rounds were not documented in the EHR 2. The attending documented stated ventilator weaning goals 81% of the time and stated sedation weaning goals 49% of the time. 3. If a stated goal was documented, there was a goal-related action approx. 83% of the time. If a stated goal was not documented, there was a goal related action 17% of the time.	NS	NS	N
Vats et al. (2010)	NS	NS	1. Variation in time spent per patient, though rounding time not significantly correlated with patient acuity 2. Rounding variation typically a product of nonessential activities (teaching, patient assessment, and family updates)	NS	NS	Y

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
Adler-Milstein et al. (2011)	1. residents felt the team was more bounded than nurses 2. Perceived communication accuracy, openness, and timeliness was similar for nurses and residents 3. Nurses perceived satisfaction with communication was higher than residents 4. Residents perceived greater collaboration and planning among the team than did nurses 5. Communication timeliness and accuracy were negatively correlated 6. Collaboration was positively correlated with communication openness, timeliness, and coordinated planning.	NS	NS	1. Nurses perceived their work to be more autonomous than residents	NS	N
Allan et al., (2010)	1. Participants felt they were better prepared to lead future resuscitation events 2. Participants indicated they were	NS	1. Participants felt they were better prepared to participate in future resuscitation events	1. Participants felt less anxious and more confidence in participating in future code events	NS	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
	more likely to notify the team leader if the resuscitation event was not being managed appropriately					
LeBlanc et al. (2010)	1. Team commitment explains the relationship between efficacy beliefs and collaborative practice 2. Team commitment is positively related to future efficacy beliefs 3. Collaborative practice is positively related to future beliefs of team commitment	NS	NS	NS	NS	N
Kim et al. (2010)	NS	1. Odds of death ratio was reduced in hospitals with that provide multidisciplinary care, especially with there is a mandatory consult or primary intensivist management)	NS	NS	NS	Y
Johnson et al. (2009)	NS	1. Decrease from 83 VAPs in 2414 vent days (34.4 VAPs per 1000 vent days) to 49 VAPs in 2094 vent days (23.4 Vaps per 1000 vent days)	1. Compliance with VAP bundle improved from approx. 50% to 94% after the implementation of MDRs	NS	NS	No Limitations Specified

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
Pronovost et al. (2008)	1. Teamwork climate increased by 10 points or more in 19 ICUs and decreased by 10 points or more in 6 ICUs. For the 6 ICUs that decreased in teamwork climate scores, nurse=physician teamwork, conflict resolution, and nurse input were rated as lower than the 10 point increase group 2. Overall, teamwork climate scores increased from approx. 47% prior to CUSP to approx. 51% a year after implementation	NS	1. Chlorhexadine was routinely available in ICU central line kits in only 20% of the 72 hospitals before CUSP. Following a letter to hospital CEOs that requested Chlorhexadine, it was observed to be stocked in 77% of hospitals and 60% of hospitals also had chlorhexadine central line kits.	NS	NS	N
Stockwell et al. (2007)	NS	NS	1. Physician management index scores were positively correlated with goal accomplishment 2. Physician management index scores and the length of rounds per patient were negative correlated	1. Attendings with more experience, years practicing since there fellowship and that were older were rated more favorably on the Physician Management Index	NS	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
Phipps and Thomas (2007)	1. The goal sheet demonstrated a positive influence of the goal sheet on perceived communication, including 73% perceiving improved communication of between nurses on different shifts 2. There was an increased perception that the unit staff worked as a team following the implementation of the daily goal sheet	1. 85% of nurses perceived the use of the daily goal sheet to have improved the care of children admitted to the PICU	NS	n's	NS	N
Pronovost et al. (2006)	NS	1. 32% (670) of contributing factors were team factors 2. 19% (386) involved verbal or written communication during routine care 3. 12% (249) involved verbal or written communication during a handoff 4. 7% (138) involved team structure and leadership	NS	NS	NS	Y
Huang et al. (2007)	1. Nurse directors tended to overestimate the team climate in their unit,	NS	NS	NS	1. Nurse directors overestimated the SAQ factor	Y

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
	but this was not significant. 2. Nurses tended to rate teamwork climate lower than physicians, though this relationship was not significant				scores of their personnel, but this was not statistically significant.	
Jain et al. (2006)	NS	1. VAP rates decreased from 7.5 per 1000 line days before the intervention to 3.2 per 1000 line days 2. The rate of urinary tract infection was reduced from 3.8 per 1000 catheter days to 2.4 per 1000 catheter days 3. BSI rates were reduced from 5.9 per 1000 line days to 3.1 per 1000 line days	NS	NS	1. The cost of ICU episode was reduced from $3406 before the intervention to $2973 following the intervention, a cost reduced of about 21%	N
Narasimhan et al. (2006)	1. Nurses and physicians perception of communication improved following the intervention and remained high 6 months after the intervention	1. The mean length of stay was shortened from 6.4 days before the intervention to 4.3 days following the intervention	1. The understanding of care goals among nurses and physicians was improved following the implementation of the goal sheet, and scores remained high 9-months after the intervention	NS	NS	Y
Sluiter et al. (2005)	1. 62% of participants indicated increased perceptions of team communication 2. Satisfaction with communication	NS	NS	1. Reduced issues associated with emotional exhaustion	NS	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
	increased from 76% before the intervention to 92% following the intervention					
Boyle and Kochinda (2004)	1. Collaborative communication scores increased among nurse and physician leaders 2. Work group cohesion increased following the intervention, but was not significant	NS	1. Perceived technical quality of care and ability to meet family needs among nurses and physicians increased after the intervention, but was not significant	1. Personal stress decreased among nurses following the intervention 2. Job satisfaction increased among nurses, but this finding was not significant 3. Intent to stay decreased among nurses following the intervention	NS	N
Wheelan et al. (2003)	NS	NS	NS	NS	NS	No Limitations Specified
Dodek and Raboud (2003)	NS	NS	1. Attending more like to be present after explicit approach introduced (85% to 93%) 2. Increased perception that there was a long-term plan of care in place for each patient following intervention (54% to 76%) 3. Increased perception that teaching time was structured during rounds (30% to 46%) 4. Increased satisfaction with the process and	NS	NS	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
			outcomes of rounds (86% to 945%) 5. Fewer residents examined there patients before rounds after the intervention than before (88% to 76%)			
Studdert et al. (2003)	Types of team conflicts: 1. conflicts centering on LST (7%) 2. Poor communication (17%) 3. Lack of leadership (9%) 4. Lack of coordination (7%) 5. Medical management (55%) 6. Belief among nurses that they were excluded from decisions (9%)	NS	NS	NS	NS	Y
Pronovost et al. (2003)	NS	1. ICU LOS decreased from 2.2 days on average per patient before the intervention to 1.1 days on average per patient after the intervention	1. Percent of residents and nurses who understood daily goals of a patient increased from 10% before the intervention to over 95% following the intervention	NS	NS	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
Pollack and Koch (2003)	NS	1. Lower incidence of PIVH/PVL was associated with high scores of leadership, conflict resolution, and coordination. Higher scores of communication and job satisfaction were also associated with lower PIVH/PVL, but were not significant. 1. Lower mortality rate was associated with higher scores for composite scores of RTs 2. Lower incidence of PIVH/PVL were associated with higher nurse composite scores	NS	NS	NS	N
Thomas et al. (2003)	1. The quality of collaboration and communication with nurses was rated high or very high by 71% by of nurses , while just 33% of nurses rated the quality of collaboration and communication with physicians as higher or very high. 2. The quality of communication and collaboration with physicians was rated	NS	NS	NS	NS	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
	as high or very high by 70% of physicians. 3. Perceptions of team climate were different based on roles, especially for difficulty speaking up, decision-making input, physician-nurse collaboration, and nurse input 4. Nurses were less satisfied with team climate than physicians					
Miller (2001)	1. Nurses and physicians had different perceptions of collaboration 2. Perceptions of communication openness between nurses and physicians, communication timeliness and satisfaction, problem solving between nurses and physicians, and problem solving within their group were rated higher by physicians than	NS	1. Physicians rated the technical quality of care higher than nurses 2. There were no observed differences in care giver interaction among personnel who attended multidisciplinary meetings	1. Physicians rated physician expertise higher than nurses rated physician expertise	NS	No Limitations Specified

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
	nurses 3. The perception of communication openness was rated as higher for day shift nurses compared to night shift nurses 4. Nurses with more experience rated perceptions of communication openness and problem solving with other nurses higher than nurses with less experience 5. Perceptions of physician leadership and communication openness were rated higher by specialty care physicians than primary care physicians					

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
Dimick et al. (2001)	NS	Patients undergoing esophageal resection: 1. Hospitals that conducted daily rounds with an ICU physician had a lower in-hospital mortality rate (4%) than hospitals that did not (approx. 14%) 2. After a patient received an esophagectomy, hospitals that did not have daily rounds by an ICU physician experienced a three-fold increase in mortality rate after adjusting for severity of illness and demographic factors 3. Patients in hospitals without daily rounds by an ICU physician was increased by 73%, or 7 days (median) 4. Hospital costs for patients undergoing esophageal resection in hospitals without ICU physician leading daily rounds was $23,335 compared to 14,424 when there was daily rounds by an ICU physician. This accounted for an increase in 61% of costs or $8,839 (median)	NS	NS	NS	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
		5. Increased risk of pulmonary insufficient, renal failure, aspiration, reintubation, and surgical complications was positively associated with hospitals that did not have daily rounds with ICU physicians				
Henneman et al. (2001)	NS	1. 40% of patients were successfully weaned off the ventilator before the intervention compared to 50% after the intervention 2. Median length of time patients received mechanical ventilation was longer before the intervention (approx. 12 days) than after (9 days). 3. Median LOS in the ICU was longer before the intervention (approx. 13 days) than after the intervention (9 days) 4. The average hospital cost was on average greater	NS	NS	NS	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
		before the intervention ($52,789; median $37,920) than after the intervention (($42,213; median $26,559)				
Baggs et al. (1999)	1. Nurses perceptions of collaboration were not associated with resident perceptions of collaboration (MICU)	1. After controlling for disease severity, nurses perceptions of collaboration predicted positive patient outcomes (MICU). Nurses at other sites and physicians/residents perceptions of collaboration across all sites were not associated with patient outcomes	NS	NS	1. There was a positive relationship between collaboration at the unit level was associated with positive patient outcomes	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
Donchin et al. (1995)	1. When verbal communications were coded, they represented 9% of all activities. 2% of these activities were between Drs and nurses and the rest were solely among Drs or solely among nurses	1. 37% of error reports noted verbal communication between nurses and physicians. (note: only 2% of activities were between nurses and physicians).	1. Activities performed by a single physician occurred 4.7% of the time 2. Activities performed by two or more physicians occurred 2.2% of the time 3. Activities performed by a single nurse occurred 84% of the time 4. Activities performed by two nurses occurred 2.7% of the time 5. Activities involving both nurses and physicians occurred 3% of the time 6. Patients daily activities were recorded only in their bedside flow sheet 47% of the time, only the physician's order sheet 7% of the time, both forms 18% of the time, and neither form 18% of the time 7. Nurse errors tended to peak 1-hr after physician rounds and around the time of their shift change	NS	NS	No Limitations Specified
Shortell et al. (1994)	NS	1. No association between care giver interaction and mortality 2. Patient LOS was associated with caregiver interaction	1. Perceptions of technical quality of care was associated with caregiver interaction 2. Perceptions of staff to meet family member needs was associated with caregiver interaction	1. Nurse turnover was negatively associated with caregiver interaction	NS	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
Baggs and Ryan (1990)	NS	NS	NS	1. Nurse perception of nurse-physician collaboration during decisions to transfer patient were associated with nurse satisfaction	NS	Y
Ilan et al. (2012)	NS	NS	NS	1. Handover duration varied significantly among physicians	NS	Y
Abraham et al (2012)	NS	NS	1. About half of communication events during handoffs were accepted without discussion 2. one third of communication events that required additional information were resolved once that info was provided 3. 4% of communication events were rejected, which resulted in a decision-making cycle 4. Collaborative problem solving was required for 11% of observed communication events that were not immediately resolved.	NS	NS	Y

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
Alvarez and Coiera (2005)	NS	NS	1. Communication events comprised 75% of round time and time spent communicated varied by role 2. Channels for communication were face-to-face (97%) and over the phone 3. Turn taking interruptions occurred in about 5.3% of communication events, and were mostly initiated by Drs (58%) 4. Conversation initiated interruptions occurred 37% of communication event time	NS	NS	No Limitations Specified
Baggs and Schmitt (1997)	1. Antecedents to collaboration involved being available and receptive	1. Improved patient care was considered by nurses and residents as a product of collaboration	NS	1. Increased job satisfaction was discussed as an outcome of working together 2. One nurse mentioned that increased nurse retention was an outcome of collaboration	NS	Y
Collins et al (2012)	NS	NS	1. Overlap in handoff content from physicians and nurses was observed. 2. Handoffs tended to by discipline specific 3. Semi-structured	NS	NS	Y

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
			communication tool (Kardex) often used by nurses during handoff as well as the nurse handoff sheet (typically used in conjunction with one another.			
Collins et al. (2011)	1. Although verbal communication is the most frequent way to exchange information, such a medium is subject to information loss/decay	NS	1. The most common way updates between disciplines take place is through discussions during and between rounds	NS	1. The current E HR system was inefficient	N
Custer et al. (2012)	1. Disjointed communication was identified as a barrier to patient care. Most communication occurs during unplanned discussions, even though rounds present a formal time for the entire team to communicate	NS	NS	NS	NS	Y
Hawryluck et al. (2002)	1. authority, education, patient needs, knowledge, resources, and time influence collaboration	NS	NS	NS	NS	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
Jirapaet et al. (2006)	1. Infective communication was considered to be a contributor to errors in nursing practice. Unclear handwriting/telephon e orders, using non standardized abbreviations contribute to ineffective communication	NS	1. Handoff procedures were considered to be a barrier to safe practice	NS	1. Insufficient staffing was considered to be a barrier to safe practice	No Limitations Specified
Lingard et al. (2004)	1. Collaboration was influenced by individuals perception of ownership of commodities (e.g., specialized knowledge and equipment) and the process of trade (e.g., allocating valued commodities)	NS	NS	NS	NS	N
Ohlinger et al. (2003)	1. Best practices for NICU Culture of Collaboration: -Communication among and between teams -Leading by example -Exhibit trust and respect Adherence to	NS	NS	NS	NS	No Limitations Specified

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
	standards of excellence/conduct -Nurture team members -Encourage conflict management					
Reader et al. (2011)	1. Leadership behaviors center on information gathering, planning and decision making, and managing team members/materials (functional leadership) 2. Team development behaviors include providing team direction and support, establishing norms, and coaching 3. The majority of cited leadership behaviors for a routine day involved managing team members (48%) 4. Most referenced team development behaviors involved the establishment of norms (41%)	NS	NS	NS	NS	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
Rosengren et al. (2007)	1. Nursing leadership was typically described in terms of availability, presence, providing acknowledgement, and facilitating care at both the individual and team level. Each of these are multifactor categories	NS	NS	NS	NS	No Limitations Specified
Heffner et al. (1996)	NS	NS	1. DNR orders were misclassified 20% of the time by nurses as full or partial during period 1 and misclassified 14% during period 2 (no real improvement). 2. Following the intervention, agreement regarding full or partial DNR status between residents and attendings increased from moderate to near perfect 3. 100% of DNR orders were descriptive following the intervention, compared to approx. 46% before the intervention 4. Approx. 69% of DNR orders were considered complete following the intervention than before the intervention (approx. 48%)	NS	NS	No Limitations Specified

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
King and Lee (1994)	1. Physicians perceptions of collaborative behavior was more favorable than nurses, but differences in perceived collaborative practice were not significant between the two groups	NS	NS	NS	NS	N
Wayne et al. (2008)	NS	NS	1. After the implementation of the new handoff system, perceived accuracy of handoff information improved.	NS	NS	N
Stockwell et al. (2005)	NS	NS	NS	NS	NS	N
Weller et al. (2011)	1. Performance was improved during measurement periods after the training was introduced when compared to before, especially when teams were led by specialists	NS	NS	NS	NS	Y
Sneve et al. (2008)	NS	1. No difference in LOS 2. Following the intervention, all DVs showed improvement. [also, prior to MDT, weight at beginning of enteral feeding was less	NS	NS	NS	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
		than weight after implementation				
DuBose et al. (2010)	NS	1. Following the implementation of QRC, there was a decrease in VAP rates (12.41 to 8.74 per 1000 ventilator days) 2. There was a lower incidence of VAP when there was full compliance (3.5% , 5.29 per 1000 vent days) than partial compliance (13.4%; 9.29 per 1000 vent days) 3. There was also a reduction in the amount of time patients spent on mechanical ventilation when there was full compliance	1. Increased compliance 3, 6, and 13 months after implementation of QRC	NS	NS	Y
Wright et al. (1996)	1. 50% of nurses thought there was inadequate communication after 6months, compared to 76% before	NS	NS	1. Less nurses felt intimated when the second questionnaire was administered 6months 2. 65% percent of nurses felt self-conscious during round presentations after 6months compared to 72%	NS	N

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
				before		
Boos et al. (2010)	NS	NS	1. CCRs provide a forum for multiple providers to reach consensus on issues relating to social challenges, complicated needs, risks, and long-term prognoses.	NS	NS	No Limitations Specified
Rangachari et al. (2010)	1. Communication among professional disciplines does not occur regularly	1. Observed CLSBSI rate of 2.5 per 1000 central line days	1. Adherence or documentation of optimal catheter site selection, skin antisepsis, and sterile barrier protocols was consistently lacking 2. Central line bundle scores were 0	NS	NS	No Limitations Specified
Manias and Street (2001)	NS	NS	1. Physicians typically start rounds before nurses have a chance to get there and that nurses could only contribute during certain portions of the discussion	1. Nurses generally feel marginalized by physicians during rounds	NS	Y
Manias and Street (2000)	1. Open communication was stymied during a global handover because the structure only allowed for the nurse coordinator to speak	NS	1. The level of specific patient information was limited which led one participant to express frustration	NS	NS	No Limitations Specified

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
Collins et al. (2010)	NS	NS	1. The preferred method of information exchange was verbal communication because clinicians felt that paper/electronic documentation was not up to date or that it was inefficient to access that information	NS	1. No formal mechanism outside of rounds where changes in nurse-physician plans/patient goals are communicated	Y
Ho et al. (2007)	1. Verbal communication was the predominant form of information exchange	NS	1. There is a lot of time pressure required to complete rounds 2. Information that is communicated comes from a variety of sources. If information became difficult to query/access, the discussion moved on to the next topic 3. Time discussing each patient during rounds was about 15-20min 4. Not all information is needed by all rounding members, which can lead some members to disassociate with the process	NS	1. Space constraints limit mobility and line of sight during rounds	N
Pickering et al. (2009)	NS	NS	1. There was a significant difference in handover scores between phase I and phase II 2. Handover scores were positively associated with clinical intention scores	NS	NS	No Limitations Specified

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
Miller et al. (2009)	NS	NS	1. Care coordination does not unfold sequentially 2. 34% of all coordination activities involved nurse-nurse discussions 3. Handovers were retrospectively focused on patient care during the previous shift, and informal conversations were used for updates on the patients current situation (nurse-nurse conversations)	NS	NS	Y
Miller et al. (2009)	NS	NS	1. Nurses involved in uni disciplinary ICU handovers did not discuss goals in reference to data and information while nurses interdisciplinary ICUs did discuss goals as well 2. Expectations and goals were discussed more frequently in interdisciplinary rounds than compared to unidisciplinary rounds	NS	NS	Y
O'Connor et al. (2009)	1. 92% of participants indicated wireless email improved speed and reliability of communication 2. 88% of participants indicated wireless email improved coordination	1. Participants perceived wireless communication to result in faster (90%) and safer (75%) patient care	NS	75% of participants indicated wireless email to reduce staff frustration	NS	Y

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
Nunnink et a. (2009)	NS	NS	1. Both training groups experienced significant improvement in objective and subjective domains	1. Confidence scores were more improved for the SBT group than video group	NS	N
Cardarelli et al. (2009)	NS	NS	1. The median rounding time per patient was 15 minutes (range 5-29) and about 26% of the time involved the presentation of patient information	NS	1. The average (salary) cost of rounds per patient was estimated to be $140.87 (range of $32.40 to $286.00 depending on role)	N
Vivian et al. (2009)	1. Trust was linked with cooperation, respect competence and professional conduct	NS	NS	1. Relationships among nurses was not described as positive	NS	No Limitations Specified
Piquette et al. (2009)	1. The need for collaboration seemed to be knowledge-specific 2. Interpersonal conflicts can arise 3. Physicians lead the team during crises 4. There is intraprofessional coordination during crises	NS	1. There is a quick transition from pre-crisis to crisis periods	1. Sometimes it is hard to escape emotional connection to patient	NS	Y
Linton et al. (2009)	1. Leadership themes include: leading by example, communication,	NS	NS	NS	NS	Y

Citation	Team Outcome	Patient Outcome	Task Outcome	Individual Outcome	Unit/ Organization Outcome	Was Limited Generalizability Specified in Study Limitations?
	ability to think outside the management square, knowing your staff, and stepping up in times of crisis					

APPENDIX C: CRITICAL INCIDENT INTERVIEW PROTOCOL

Section 1: Teamwork in General

(These questions will define the scope of the interview and help determine what examples are selected for the critical incident technique described in section 2. Specifically, the goal of section 1 is to identify what aspects of teamwork they think are most important and when they are important. These questions are intended to be supplemented by additional questions listed in section 2).

First, we would like to start by asking a few questions to guide the remainder of our conversation.

1) What does a typical day at work look like [in your role]? Please walk me through your day and tell me about the tasks you perform and the people you interact with.
   a) How does the set of people change depending on the activities you are performing that day?
   b) Who, in terms of position rather than name, do you interact with most often in your role?
   c) If you were to define a 'team' of people you work with, which of these individuals do you consider as part of your team?
      i) Do you consider patients to be a part of your team?

2) If you were to define the most important components of teamwork, what would they be and why?

   Provide interviewees with a copy of the team performance framework (Figure 1) and describe its components. For example, say "Teamwork is often defined in terms of inputs, processes, and outcomes. Here, we see how inputs of teamwork, which include things like team member composition and task characteristics, are transformed into team outputs such as efficiency and safety by things such as communication, planning, and coordination."

   a) What do you think are the most important teamwork processes for the activities you perform?
   b) During which tasks is teamwork most important?

Additional probes, if appropriate:

- In a few words, how would you summarize the general role of teamwork in the ICU?
- In your experience, what are the skills needed to work effectively as a team?

Section 2: Critical Incident Probes

(First, ensure interviewees have specified at least one team task/activity and/or team competency/process as being important. The goal is to dive deeper from examples provided in section 1. The focus will be on eliciting information about one positive and one negative example of the event and/or competency/process).

2A. Questions Targeting Team Competencies/Processes

(If the discussion in section 1 focused primarily on team competencies, rather than tasks, start here. Otherwise, start at section 2B and skip 2A).

3) You mentioned that [insert competency/process] was an important component of teamwork. Thinking back on your experiences, could you describe an example of a task or event that happened when [insert competency/process] was effective and when [insert competency/process] was not effective or needed improvement?
   a) Can you provide a general timeline of activities, key decision-points, or circumstances involved *before* [insert task/event]
   b) Can you provide a general timeline of activities, key decision-points, or circumstances involved *during* [insert task/event]
   c) Can you provide a general timeline of activities, key decision-points, or circumstances involved *after* [insert task/event]

Note: As first about a positive (or negative) example, then follow that example with the alternative. That is, do not ask the participant to simultaneously describe both a positive and negative example.

Additional probes, if appropriate:
(First, take notes on *what* happened in the examples, then probe for additional information that help explain *why* the events happened. Emphasize key decision-making processes in bold).

<u>Decisions</u>:
- What were your goals?
- What options were you working with? How did you know which one was right?

<u>Situation assessment</u>:
- What information did you have or need?
- What cues were you attending to?
- What were your expectations?
- What did you think was happening then?

<u>Knowledge</u>:
- What experience helped you in the situation?
  i) Did you seek out any guidance?

<u>Composition:</u>
- What team members were involved in the task/event and what was their role? How did they help/hinder task execution?

4) Comparisons and Lessons Learned
   a) How did the actions taken in the 'good' example differ from those taken in the example in which you felt there was a need for improvement?
      i) What would you like to do differently?
      ii) What tools (e.g., protocols, training) or technology would have made a difference?
   b) What would you say are key lessons that were learned from these events?

2B. Questions Targeting Team Tasks or Events

5) You mentioned that [insert event/activity] was an important task that depended on teamwork. Could you please tell me about a time when [insert event/activity] was handled well and when [insert event/activity] did not go as well as you had hoped?
   a) Can you provide a general timeline of activities, key decision-points, or circumstances involved *before* [insert task/event]
   b) Can you provide a general timeline of activities, key decision-points, or circumstances involved *during* [insert task/event]
   c) Can you provide a general timeline of activities, key decision-points, or circumstances involved *after* [insert task/event]

Note: As first about a positive (or negative) example, then follow that example with the alternative. That is, do not ask the participant to simultaneously describe both a positive and negative example.

Additional probes, if appropriate:
(First, take notes on *what* happened in the examples, then probe for additional information that help explain *why* the events happened. Emphasize key decision-making processes in bold).

Decisions:
- What were your goals?
- What options were you working with? How did you know which one was right?

Situation assessment:
- What information did you have or need?
- What cues were you attending to?
- What were your expectations?
- What did you think was happening then?

Knowledge:
- What experience helped you in the situation?
   i) Did you seek out any guidance?

Composition:
- What team members were involved in the task/event and what was their role? How did they help/hinder task execution?

Team Factors:
- Was there anything especially relevant to team factors? (Use probes below if necessary)
   - E.g., coordination, communication, cooperation

6) Comparisons and Lessons Learned
   a) How did the actions taken in the 'good' example differ from those taken in the example in which you felt there was a need for improvement
      i) What would you like to do differently?
      ii) What tools (e.g., protocols, training) or technology would have made a difference?
   b) What would you say are key lessons that were learned from these events?

Section 3: Wrap-up Questions

7) What are the three most significant challenges (e.g., barriers/disruptors) to effective teamwork in your unit?
   a) What are the three most significant challenges to effective [insert competency/process from section 2]?

8) What are the three most significant facilitators to effective teamwork in your unit?
   a) What are the three most significant facilitators to effective [insert competency/process from section 2]?

9) If you could suggest three mechanisms that would improve teamwork, what would they be and why?

Section 4: Conclusions

Is there anything else you would like to share/discuss with us?

Ok, great. Thank you for taking the time to speak with us today, we greatly appreciate it. Please feel free to call us back at ___-___-______ or send us an email at ____@_____ if you would like to add anything or have any additional questions.

APPENDIX D: ICU BEHAVIORAL MARKER SYTEM

Observation Type:____________________ Duration (min):________ Rater Initials:_________________ Date:_______________

Team Size:____________ Size Variability: ☐ Dynamic ☐ Stable

Team Diversity: ☐ Nurses Only ☐ Physicians Only ☐ PT Only ☐ Multidisciplinary

Dimension	Sub-Dimension	*Rating	Observation Notes	*Dimension Rating and Debrief Notes
Communication	Style			
	Content			
	Closed-Loop			
Leadership	Delegation			
	Norms			
Backup and Supporting Behavior	Offering Backup/Support			
	Seeking Backup/Support			
	Feedback			
Team Decision-Making	Goals			
	Contingency Planning			
	Updating and Revising			

***1 (Poor):** Performance was expected, but not observed; Performance consistently demonstrated negative teamwork behaviors.
2 (Marginal)
3 (Neutral/Acceptable): Performance was adequate. Team members demonstrated positive teamwork behaviors, but also showed areas for improvement; Team competency acknowledged, but opportunities to further demonstrate competency precluded due to patient conditions or situation.
4 (Good)
5 (Very Effective): Performance consistently demonstrated positive teamwork behaviors throughout the entire observation.

Communication - Communication refers to the style and structure of how information is conveyed between team members. Communication entails exchanging messages using standardized protocols with appropriate terminology in a manner that is clear, accurate, and succinct. A key feature of communication exchanges is that they are closed-loop; the sender conveys information, the receiver confirms the receipt of information, and the sender clarifies any misunderstandings.

Style: Messages are conveyed in a manner that is clear and succinct.

Poor	Good
• Uses technical jargon when discussing care plan with patients and/or family members. • Multiple speakers presenting information simultaneously. • Volume is too low and pace is fast. • Shouting between team members. • Vague / indirect communication • Verbose communication. • Communication interrupted/disrupted.	• Uses lay terms when discussing care plan with patients and/or family members. • Only one speaker presenting information at a time. • The volume of speech is appropriate for all team members to hear. • Communication is calm, clear, and explicit. • Manages interruptions/disruptions appropriately.

Content: Messages are conveyed with appropriate structure and accuracy.

Poor	Good
• Standard communication protocols/tools are not used/followed. • Big picture situational summaries not provided.	• Appropriate communication protocols/tools are used/followed. • Big picture summaries are provided. • Rationales for orders and task assignments conveyed.

Closed-Loop: The sender conveys information, the receiver confirms the receipt of information, and the sender clarifies any misunderstandings.

Poor	Good
• Directives carried out without confirming intent. • Receipt of information is not confirmed. • Messages are sent electronically without subsequent face-to-face communication.	• Directive confirmed and intent to execute verbalized. • Receipt of communication acknowledged for both face-to-face and electronic communication. • Electronic delivery of messages is followed-up with face-to-face communication.

Leadership - Team leadership refers to the management of team resources/personnel, establishment of team norms, and provision of opportunities to foster the development of knowledge and skills. Team leaders ensure there is clarity of team member roles/responsibilities and that input from all team members is welcomed.

<u>Delegation:</u> The management of team resources/personnel.

Poor	Good
• Team member asks for clarification with no resolution. • Care plans and responsibilities are dictated without input from other team members. • Workload is arbitrarily assigned to clinical team members. • Expectations of taskwork assignments are not established.	• Roles and responsibilities delegated clearly • Roles and responsibilities assumed implicitly with clear coordination and synchronization. • Leader confirms team has a shared understanding of care plans and priorities. • Team leader describes the importance of assigned taskwork in relation to care goals.

<u>Norms:</u> The establishment of standards and models of behavioral expectations.

Poor	Good
• Input from team members is dismissed or discouraged based on role and status hierarchies. • Good work is not acknowledged.	• Team establishes an inclusive atmosphere by seeking input from all team members and encouraging questions, regardless of role (including the patient). • Team leader acknowledges good work and provides positive reinforcement. • New team members introduce themselves to the clinical team.

Backup and Supportive Behavior - Backup and supportive behavior refers to proactively seeking and providing task-related assistance.

Offering Backup/Support: Offering task-related assistance.

Poor	Good
• Does not offer assistance when another team member is overloaded. • Do not support each other's decisions in front of patient's and family members. • Team members do not cross check to confirm recommended plans are being executed.	• Reallocates work when a more critical task is presented. • Offers help throughout the shift/performance episode. • Team members support each other's decisions in front of patients and family members.

Seeking Backup/Support: Proactively requesting task-related assistance.

Poor	Good
• The page system is used to solicit assistance for planned clinical activities. • Does not seek assistance during emergent event. • Requests assistance from overloaded team member. •	• Informs other team members when assistance is needed prior to planned clinical activities. • Immediately requests assistance during acute situation. • Recognizes when overloaded and engages appropriate resources

Feedback: The provision of error correction and developmental behaviors.

Poor	Good
• Sr. clinician intervenes without explaining rationale. • Assistance and feedback not provided during unfamiliar tasks. • Team member receives no feedback when errors or near misses occur.	• Identifies errors/near misses and assists with remediation. • Assistance and feedback are provided for unfamiliar tasks. • Provides feedback when errors or near misses occur. • Teaching opportunities are provided through probes for additional information or by offering additional information about the case or treatment plan. • Verbalizes discrepancies.

Team Decision-Making - Team decision-making refers to the team's ability to determine goals, develop plans and strategies for task accomplishment, and identify contingencies.

Planning and Establishing Goals: Team members identify care goals, methods to achieve goals, and anticipated outcomes (prospective).

Poor	Good
• Treatment plans are executed without a formal discussion. • Anticipated outcomes of treatment activities are not identified. • Treatment plans developed without diverse input.	• Team members deliberately discuss, propose, and prioritize the planned course of patient care for each patient. • Team members define anticipated outcomes. • Team members discus resource needs to accomplish goals

Contingency Planning: Team members prepare for likely scenarios that alter care plans.

Poor	Good
• Does not consider unanticipated outcomes or barriers/challenges that may impede progress. • Team members do not specify alternate treatment plans should unexpected event occur. • Alternate plans are specified without justification.	• Identifies conditions or events that may alter treatment plans, including barriers and challenges that may impede progress. • Specifies alternative courses of action for treatment plans. • Discuss why there is a need for alternate treatment plans.

Updating and Revising: Team members discuss updates and make revisions to care goals as needed (retrospective).

Poor	Good
• Treatment plans are not modified in response to changing patient conditions. • Team members do not discuss the underlying factors that prompted care plans to change. • Unique information not shared. • Assessment of care plan effectiveness not shared among team members.	• Review information relating to care, whether those goals have been achieved, and what needs to be accomplished if those goals have not been realized. • Identify any challenges encountered while executing care plans and emerging issues. • Relevant team members (including P/F) are informed of updates to care goals and pans, changing patient conditions, and following consults with inter-unit staff.

APPENDIX E: UNIVERSITY OF CENTRAL FLORIDA IRB APPROVAL

University of Central Florida

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Exempt Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Aaron S. Dietz**

Date: **August 01, 2014**

Dear Researcher:

On 8/1/2014, the IRB approved the following activity as human participant research that is exempt from regulation:

Type of Review:	Exempt Determination
Project Title:	Validation of an ICU Teamwork Measure
Investigator:	Aaron S Dietz
IRB Number:	SBE-14-10476
Funding Agency:	Gordon and Betty Moore Foundation(GBMF)
Grant Title:	Intensive Care Unit (ICU)/Acute Care Multi-site Demonstration Project (Grant #3186.01)
Research ID:	N/A

This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these changes affect the exempt status of the human research, please contact the IRB. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Signature applied by Joanne Muratori on 08/01/2014 02:31:30 PM EDT

Joanne Muratori

IRB Coordinator

Page 1 of 1

APPENDIX F: JOHNS HOPKINS MEDICINE IRB APPROVAL

Office of Human Subjects Research
Institutional Review Boards

1620 McElderry Street, Reed Hall, Suite B-130
Baltimore, Maryland 21205-1911
410-955-3008
410-955-4367 Fax
e-mail: jhmirb@jhmi.edu

Date: July 28, 2014

ACKNOWLEDGEMENT

Review Type:	Not Human Subjects Research (NHSR)
PI Name:	Michael Rosen
Study #:	IRB00034010
Study Name:	Validation of an ICU Teamwork Measure
Committee Chair:	Susan Bassett
Committee:	IRB-X

Date of acknowledgement: July 28, 2014

Date of expiration: July 28, 2017

The JHM IRB has determined that the above-referenced New Application does not involve human subjects research under the DHHS or FDA regulations. You may proceed with this project without further interaction with the JHM IRB. If there are changes in this project that may affect this determination, you should consult with the JHM IRB before making those changes.

The publishing of QA/QI projects does not require IRB approval. The publication must clearly state this was a QA/QI project and cannot refer to it as a research study, since research studies require IRB approval before they are implemented.

To keep the JHM IRB files current, we are assigning a 3-year expiration date to projects [studies] that qualify as exempt or not human subjects research. You will receive an email notification prior to the expiration date, allowing you to extend this project [the research] by completing an 'Extend Approval' activity (a continuing review application is not required).

Study documents:

HIPAA Form 4:

Supplemental Study Documents:

Additional Supplemental Study Documents:
LOS_ASapirstein.pdf
Marker System .docx
LOS_PLipsett.pdf
LOS_KEarsing.pdf

Protocol:
Markers eIRB.doc

Study Team Members:
Aaron Dietz

The JHMIRB is constituted to meet the requirements of the Privacy Rule at section 45 CFR 164.512(i)(1)(i)(B) and is authorized and qualified to serve as the Privacy Board for human subjects research applications conducted by Hopkins' faculty members. The JHM IRB reviewed your request to waive authorization the above-referenced project. The IRB determined that all specific criteria for a waiver of authorization were met, as follows:
(A) The use or disclosure of protected health information involves no more than minimal risk to the privacy of individuals, based on, at least, the presence of the following elements;
(1) An adequate plan to protect the identifiers from improper use and disclosure;
(2) An adequate plan to destroy the identifiers at the earliest opportunity consistent with conduct of the research, unless there is a health or research justification for retaining the identifiers or such retention is otherwise required by law; and
(3) Adequate written assurances that the protected health information will not be reused or disclosed to any other person or entity, except as required by law, for authorized oversight of the research study, or for other research for which the use or disclosure of protected health information would be permitted;
(B) The research could not practicably be conducted without the waiver or alteration; and
(C) the research could not practicably be conducted without access to and use of the protected health information.

The Johns Hopkins Institutions operates under multiple Federal-Wide Assurances: The Johns Hopkins University School of Medicine - FWA00005752, The Johns Hopkins University School of Nursing - FWA00006088, The Johns Hopkins Hospital and Johns Hopkins Health Systems - FWA00006087, Johns Hopkins Bayview Medical Center - FWA00006089, Howard County General Hospital - FWA00005743, Hugo W. Moser Research Institute at Kennedy Krieger, Inc. - FWA00005719, Johns Hopkins Community Physicians - FWA00002251, Suburban Hospital and Health System - FWA00005924

APPENDIX G: COPYRIGHT AGREEMENT, JOURNAL OF CRITICAL CARE

ELSEVIER LICENSE
TERMS AND CONDITIONS

Oct 29, 2014

This is a License Agreement between Aaron Dietz ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Aaron Dietz
Customer address	1200 S. Conkling St.
	BALTIMORE, MD 21224
License number	3498301431761
License date	Oct 29, 2014
Licensed content publisher	Elsevier
Licensed content publication	Journal of Critical Care
Licensed content title	A systematic review of teamwork in the intensive care unit: What do we know about teamwork, team tasks, and improvement strategies?
Licensed content author	Aaron S. Dietz,Peter J. Pronovost,Pedro Alejandro Mendez-Tellez,Rhonda Wyskiel,Jill A. Marsteller,David A. Thompson,Michael A. Rosen
Licensed content date	December 2014
Licensed content volume number	29
Licensed content issue number	6
Number of pages	7
Start Page	908
End Page	914
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes

Will you be translating?	No
Title of your thesis/dissertation	The Development and Testing of a Measurement System to Assess Intensive Care Unit Team Performance
Expected completion date	Nov 2014
Estimated size (number of pages)	200
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	0.00 USD

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

For journal authors: the following clauses are applicable in addition to the above: Permission granted is limited to the author accepted manuscript version* of your paper.

***Accepted Author Manuscript (AAM) Definition:** An accepted author manuscript (AAM) is the author's version of the manuscript of an article that has been accepted for publication and which may include any author-incorporated changes suggested through the processes of submission processing, peer review, and editor-author communications. AAMs do not include other publisher value-added contributions such as copy-editing, formatting, technical enhancements and (if relevant) pagination.

You are not allowed to download and post the published journal article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version. A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx. As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier's online service ScienceDirect (www.sciencedirect.com). That e-mail will include the article's Digital Object Identifier (DOI). This number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.

Posting to a repository: Authors may post their AAM immediately to their employer's institutional repository for internal use only and may make their manuscript publically available after the journal-specific embargo period has ended.

Please also refer to Elsevier's Article Posting Policy for further information.

18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only.. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.

20. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

Elsevier Open Access Terms and Conditions

Elsevier publishes Open Access articles in both its Open Access journals and via its Open Access articles option in subscription journals.

Authors publishing in an Open Access journal or who choose to make their article Open Access in an Elsevier subscription journal select one of the following Creative Commons user licenses, which define how a reader may reuse their work: Creative Commons Attribution License (CC BY), Creative Commons Attribution – Non Commercial - ShareAlike (CC BY NC SA) and Creative Commons Attribution – Non Commercial – No Derivatives (CC BY NC ND)

Terms & Conditions applicable to all Elsevier Open Access articles:

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation.

The author(s) must be appropriately credited.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

Additional Terms & Conditions applicable to each Creative Commons user license:

CC BY: You may distribute and copy the article, create extracts, abstracts, and other revised versions, adaptations or derivative works of or from an article (such as a translation), to include in a collective work (such as an anthology), to text or data mine the article, including for commercial purposes without permission from Elsevier

CC BY NC SA: For non-commercial purposes you may distribute and copy the article, create extracts, abstracts and other revised versions, adaptations or derivative works of or from an article (such as a translation), to include in a collective work (such as an anthology), to text and data mine the article and license new adaptations or creations under identical terms without permission from Elsevier

CC BY NC ND: For non-commercial purposes you may distribute and copy the article and include it in a collective work (such as an anthology), provided you do not alter or modify the article, without permission from Elsevier

Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.

Commercial reuse includes:

- Promotional purposes (advertising or marketing)
- Commercial exploitation (e.g. a product for sale or loan)
- Systematic distribution (for a fee or free of charge)

Please refer to Elsevier's Open Access Policy for further information.

21. **Other Conditions**:

v1.6

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

APPENDIX H: COPYRIGHT AGREEMENT, BMJ QUALITY & SAFETY

BMJ PUBLISHING GROUP LTD. LICENSE TERMS AND CONDITIONS

Oct 29, 2014

This is a License Agreement between Aaron Dietz ("You") and BMJ Publishing Group Ltd. ("BMJ Publishing Group Ltd.") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by BMJ Publishing Group Ltd., and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3498301052868
License date	Oct 29, 2014
Licensed content publisher	BMJ Publishing Group Ltd.
Licensed content publication	BMJ Quality and Safety
Licensed content title	A systematic review of behavioural marker systems in healthcare: what do we know about their attributes, validity and application?
Licensed content author	Aaron S Dietz, Peter J Pronovost, Kari N Benson, Pedro Alejandro Mendez-Tellez, Cynthia Dwyer, Rhonda Wyskiel, Michael A Rosen
Licensed content date	Aug 25, 2014
Type of Use	Dissertation/Thesis
Requestor type	Author of this article
Format	Print and electronic
Portion	Figure/table/extract
Number of figure/table/extracts	5
Descriptionof figure/table/extracts	Figure 1, Figure 2, Supplementary Table 1, 2, and 3. Text from the Introduction, Background, Methods, The State of Quality, and Implications for Research and Practice.
Will you be translating?	No
Circulation/distribution	2
Title of your thesis / dissertation	The Development and Testing of a Measurement System to Assess Intensive Care Unit Team Performance
Expected completion date	Nov 2014
Estimated size(pages)	200
BMJ VAT number	674738491
Billing Type	Invoice
Billing address	1200 S. Conkling St.
	Apt. 238
	BALTIMORE, MD 21224

	United States
Permissions Cost	0.00 USD
VAT (0.00%)	0.00 USD
Total	0.00 USD

Terms and Conditions

BMJ Group Terms and Conditions for Permissions

When you submit your orderyou are subject to the terms and conditions set out below. You will also haveagreed to the Copyright Clearance Center's ("CCC") terms and conditionsregarding billing and payment https://s100.copyright.com/App/PaymentTermsAndConditions.jsp. CCC are acting as the BMJ Publishing Group Limited's ("BMJ Group's") agent.

Subject to the terms set outherein, the BMJ Group hereby grants to you (the Licensee) a non-exclusive,non-transferable licence to re-use material as detailed in your request forthis/those purpose(s) only and in accordance with the following conditions:

1) **Scope of Licence:** Useof the Licensed Material(s) is restricted to the ways specified by you duringthe order process and any additional use(s) outside of those specified in thatrequest, require a further grant of permission.

2) **Acknowledgement:** Inall cases, due acknowledgement to the original publication with permission fromthe BMJ Group should be stated adjacent to the reproduced Licensed Material.The format of such acknowledgement should read as follows:

"Reproduced from [publicationtitle, author(s), volume number, page numbers, copyright notice year] withpermission from BMJ Publishing Group Ltd."

3) **Third Party Material**:BMJ Group acknowledges to the best of its knowledge, it has the rights to licenceyour reuse of the Licensed Material, subject always to the caveat thatimages/diagrams, tables and other illustrative material included within, whichhave a separate copyright notice, are presumed as excluded from the licence. Therefore,you should ensure that the Licensed Material you are requesting is original toBMJ Group and does not carry the copyright of another entity (as credited inthe published version). If the credit line on any part of the material you haverequested in any way indicates that it was reprinted or adapted by BMJ Groupwith permission from another source, then you should seek permission from thatsource directly to re-use the Licensed Material, as this is outside of thelicence granted herein.

4) **Altering/ModifyingMaterial:** The text of any material for which a licence is granted may notbe altered in any way without the prior express permission of the BMJ Group. Subject to Clause 3 above however, single figure adaptations do not require BMJGroup's approval; however, the adaptation should be credited as follows:

"Adapted by permission fromBMJ Publishing Group Limited. [publication title, author, volume number, pagenumbers, copyright notice year]

5) **Reservation of Rights:**The BMJ Group reserves all rights not specifically granted in the combinationof (i) the licence details provided by you and accepted in the course of thislicensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment Terms and Conditions.

6) **Timing of Use:** Firstuse of the Licensed Material must take place within 12 months of the grant ofpermission.

7). **Creation of Contractand Termination:** Once you have submitted an order via Rightslink and thisis received by CCC, and subject to you completing accurate details of yourproposed use, this is when a binding contract is in effect and our acceptanceoccurs. As you are ordering rights from a periodical, to the fullest extentpermitted by law, you will have no right to cancel the contract from this pointother than for BMJ Group's material breach or fraudulent misrepresentation oras otherwise permitted under a statutory right. Payment must be made inaccordance with CCC's Billing and Payment Terms and conditions. In the eventthat you breach any material condition of these terms and condition or any of CCC's Billing and Payment Terms and Conditions, the license is automatically terminated uponwritten notice from the BMJ Group or CCC or as otherwise provided for in CCC's Billing and Payment Terms and Conditions, where these apply.. Continued use of materials wherea licence has been terminated, as well as any use of the Licensed Materialsbeyond the scope of an unrevoked licence, may constitute intellectual propertyrights infringement and BMJ Group reserves the right to take any and all actionto protect its intellectual property rights in the Licensed Materials.

8. **Warranties:** BMJGroup makes no express or implied representations or warranties with respect tothe Licensed Material and to the fullest extent permitted by law this isprovided on an "as is" basis. For the avoidance of doubt BMJ Group does notwarrant that the Licensed Material is accurate or fit for any particularpurpose.

9. **Limitation ofLiability:** To the fullest extent permitted by law, the BMJ Group disclaimsall liability for any indirect, consequential or incidental damages (includingwithout limitation, damages for loss of profits, information or interruption)arising out of the use or inability to use the Licensed Material or theinability to obtain additional rights to use the Licensed Material. To thefullest extent permitted by law, the maximum aggregate liability of the BMJGroup for any claims, costs, proceedings and demands for direct losses causedby BMJ Group's breaches of its obligations herein shall be limited to twice theamount paid by you to CCC for the licence granted herein.

10. **Indemnity:** You hereby indemnify and holdharmless the BMJ Group and their respective officers, directors, employees andagents, from and against any and all claims, costs, proceeding or demands arisingout of your unauthorised use of the Licensed Material.

11. **No Transfer of License:** This licence ispersonal to you, and may not be assigned or transferred by you without priorwritten consent from the BMJ Group or its authorised agent(s). BMJ Groupmay assign or transfer any of its rights and obligations under this Agreement,upon written notice to you.

12. **No Amendment Except inWriting:** This licence may not be amended except in a writing signed by bothparties (or, in the case of BMJ Group, by CCC on the BMJ Group's behalf).

13. **Objection to Contraryterms:** BMJ Group hereby objects to any terms contained in any purchaseorder, acknowledgment, check endorsement or other writing prepared by you,which terms are inconsistent with these terms and conditions or CCC's Billing and Payment Terms and Conditions. These terms and conditions, together with CCC's Billing and Payment Terms and Conditions (which to the extent they are consistent areincorporated herein), comprise the entire agreement between you and BMJ Group (and CCC) and the Licensee concerning this licensing transaction. In the event ofany conflict between your obligations established by these terms and conditionsand those established by CCC's Billing and Payment Terms and Conditions, theseterms and conditions shall control.

14. **Revocation:** BMJGroup or CCC may, within 30 days of issuance of this licence, deny thepermissions described in this licence at their sole discretion, for any reasonor no reason, with a full refund payable to you should you have not been ableto exercise your rights in full. Notice of such denial will be made using thecontact information provided by you. Failure to receive such notice from BMJGroup or CCC will not, to the fullest extent permitted by law alter orinvalidate the denial. For the fullest extent permitted by law in no event willBMJ Group or CCC be responsible or liable for any costs, expenses or damageincurred by you as a result of a denial of your permission request, other thana refund of the amount(s) paid by you to BMJ Group and/or CCC for denied permissions.

15. **Restrictions to thelicense:**

15.1 Promotion: BMJ Group will not give permission to reproduce infull or in part any Licensed Material for use in the promotion of thefollowing:

a) non-medical products thatare harmful or potentially harmful to health: alcohol, baby milks and/or, sunbeds

b) medical products that donot have a product license granted by the Medicines and Healthcare productsRegulatory Agency (MHRA) or its international equivalents. Marketing of theproduct may start only after data sheets have been released to members of themedical profession and must conform to the marketing authorization contained inthe product license.

16. Translation: This permission is granted for non-exclusive worldEnglish language rights only unless explicitly stated in your licence. Iftranslation rights are granted, a professional translator should be employedand the content should be reproduced word for word preserving the integrity ofthe content.

17. **General:**Neither party shall be liable for failure, default or delay in performingits obligations under this Licence, caused by a Force Majeure event which shallinclude any act of God, war, or threatened war, act or threatened act ofterrorism, riot, strike, lockout, individual action, fire, flood, drought,tempest or other event beyond the reasonable control of either party.

17.1 Inthe event that any provision of this Agreement is held to be invalid, theremainder of the provisions shall continue in full force and effect.

17.2 Thereshall be no right whatsoever for any third party to enforce the terms andconditions of this Agreement. The Parties hereby expressly wish to exclude theoperation of the Contracts (Rights of Third Parties) Act 1999 and any otherlegislation which has this effect and is binding on this agreement.

17.3 Tothe fullest extent permitted by law, this Licence will be governed by the lawsof England and shall be governed and construed in accordance with the laws of England. Any action arising out of or relating to this agreement shall be brought in courtssituated in England save where it is necessary for BMJ Group for enforcement tobring proceedings to bring an action in an alternative jurisdiction.

REFERENCES

Abraham, J., Kannampallil, T. G., & Patel, V. L. (2012). Bridging gaps in handoffs: A continuity of care based approach. *Journal of Biomedical Informatics, 45*(2), 240-254. doi:10.1016/j.jbi.2011.10.011; 10.1016/j.jbi.2011.10.011

Allan, C. K., Thiagarajan, R. R., Beke, D., Imprescia, A., Kappus, L. J., Garden, A., . . . Weinstock, P. H. (2010). Simulation-based training delivered directly to the pediatric cardiac intensive care unit engenders preparedness, comfort, and decreased anxiety among multidisciplinary resuscitation teams. *The Journal of Thoracic and Cardiovascular Surgery, 140*(3), 646-652. doi:10.1016/j.jtcvs.2010.04.027; 10.1016/j.jtcvs.2010.04.027

Alvarez, G., & Coiera, E. (2005). Interruptive communication patterns in the intensive care unit ward round. *International Journal of Medical Informatics, 74*(10), 791-796. doi:10.1016/j.ijmedinf.2005.03.017

American Institutes for Research. (2010). *TeamSTEPPS teamwork perceptions questionnaire (T-TPQ) manual*. Washington, DC: American Institutes for Research.

Andersen, P. O., Jensen, M. K., Lippert, A., Ostergaard, D., & Klausen, T. W. (2010). Development of a formative assessment tool for measurement of performance in multi-professional resuscitation teams. *Resuscitation, 81*(6), 703-711. doi:10.1016/j.resuscitation.2010.01.034; 10.1016/j.resuscitation.2010.01.034

ANTS. (2012). ANTS: A behavioural marker system for rating anaesthetists' non-technical skills. Retrieved from http://www.abdn.ac.uk/iprc/ants/

Arthur, W., Woehr, D. J., & Maldegen, R. (2000). Convergent and discriminant relationships of assessment center dimensions: A conceptual and empirical reexamination of the assessment center construct-related validity paradox. *Journal of Management, 26*(4), 813-834. doi:10.1177/014920630002600410

Baggs, J. G., Norton, S. A., Schmitt, M. H., & Sellers, C. R. (2004). The dying patient in the ICU: Role of the interdisciplinary team. *Critical Care Clinics, 20*(3), 525-40, xi. doi:10.1016/j.ccc.2004.03.008

Baggs, J. G., Schmitt, M. H., Mushlin, A. I., Mitchell, P. H., Eldredge, D. H., Oakes, D., & Hutson, A. D. (1999). Association between nurse-physician collaboration and patient outcomes in three intensive care units. *Critical Care Medicine, 27*(9), 1991-1998.

Baker, D. P., Gustafson, S., Beaubien, J., Salas, E., & Barach, P. (2005). *Medical teamwork and patient safety: The evidence-based relation.* (No. 05-0053). Rockville, MD: Agency for Healthcare Research and Quality.

Baker, D. P., & Salas, E. (1992). Principles for measuring teamwork skills. *Human Factors, 34*(4), 469-475.

Baker, D. P., & Salas, E. (1997). Principles for measuring teamwork: A summary. In M. T. Brannick, E. Salas & E. Prince (Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp. 331-355). Mahwah, NJ: Lawrence Erlbaum Associates.

Boyle, D. K., & Kochinda, C. (2004). Enhancing collaborative communication of nurse and physician leadership in two intensive care units. *The Journal of Nursing Administration, 34*(2), 60-70.

Brannick, M.T., & Prince, C. (1997). An overview of team performance measurement. *Team performance assessment and measurement: Theory, methods, and applications* (pp. 3-18). Mahwah, NJ: Lawrence Erlbaum Associates.

Brennan, R. L. (2001). *Generalizability theory*. New York, NY: Springer.

Buljac-Samardzic, M., Dekker-van Doorn, C. M., van Wijngaarden, J. D., & van Wijk, K. P. (2010). Interventions to improve team effectiveness: A systematic review. *Health Policy (Amsterdam, Netherlands), 94*(3), 183-195. doi:10.1016/j.healthpol.2009.09.015; 10.1016/j.healthpol.2009.09.015

Burke, C. S., Stagl, K. C., Klein, C., Goodwin, G. F., Salas, E., & Halpin, S. M. (2006). What type of leadership behaviors are functional in teams? A meta analysis. *The Leadership Quarterly, 17*(3), 288-307. doi:10.1016/j.leaqua.2006.02.007

Burke, C. S., Salas, E., Wilson-Donnelly, K., & Priest, H. (2004). How to turn a team of experts into an expert medical team: Guidance from the aviation and military communities. *Quality & Safety in Health Care, 13 Suppl 1*, i96-104. doi:10.1136/qhc.13.suppl_1.i96

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105.

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining competencies and establishing team training requirements. In R. Guzzo, & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 333-380). San Francisco,CA: Jossey-Bass.

Capella, J., Smith, S., Philp, A., Putnam, T., Gilbert, C., Fry, W., . . . Remine, S. (2010). Teamwork training improves the clinical care of trauma patients. *Journal of Surgical Education, 67*(6), 439-443. doi:10.1016/j.jsurg.2010.06.006; 10.1016/j.jsurg.2010.06.006

Carayon, P., & Friesdorf, W. (2006). Human factors and ergonomics in medicine. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (Third Edition ed., pp. 1517-1537). Hoboken, NJ, USA.: John Wiley & Sons, Inc. doi:10.1002/0470048204.ch58

Cardinet, J., Johnson, S., & Gianreto, P. (2010). *Applying generalizability theory using EduG*. New York, NY: Routledge.

Cardinet, J., Tourneur, Y., & Allal, L. (1976). The symmetry of generalizability theory: Applications to educational measurement. *Journal of Educational Measurement, 13*(2), 119-135.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284-290.

Collins, S. A., Mamykina, L., Jordan, D., Stein, D. M., Shine, A., Reyfman, P., & Kaufman, D. (2012). In search of common ground in handoff documentation in an intensive care unit. *Journal of Biomedical Informatics, 45*(2), 307-315. doi:10.1016/j.jbi.2011.11.007; 10.1016/j.jbi.2011.11.007

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: Theory and application. *The American Journal of Medicine, 119*(2), 166.e7-166.16. doi:10.1016/j.amjmed.2005.10.036

Cooper, S., Cant, R., Porter, J., Missen, K., Sparkes, L., McConnell-Henry, T., & Endacott, R. (2012). Managing patient deterioration: Assessing teamwork and individual performance. *Emergency Medicine Journal : EMJ,* doi:10.1136/emermed-2012-201312

Crossley, J., Davies, H., Humphris, G., & Jolly, B. (2002). Generalisability: A key to unlock professional assessment. *Medical Education, 36*(10), 972-978.

Crossley, J., Marriott, J., Purdie, H., & Beard, J. D. (2011). Prospective observational study to evaluate NOTSS (non-technical skills for surgeons) for assessing trainees' non-technical performance in the operating theatre. *The British Journal of Surgery, 98*(7), 1010-1020. doi:10.1002/bjs.7478; 10.1002/bjs.7478

Crossley, J., Russell, J., Jolly, B., Ricketts, C., Roberts, C., Schuwirth, L., & Norcini, J. (2007). 'I'm pickin' up good regressions': The governance of generalisability analyses. *Medical Education, 41*(10), 926-934. doi:10.1111/j.1365-2923.2007.02843.x

Custer, J. W., White, E., Fackler, J. C., Xiao, Y., Tien, A., Lehmann, H., & Nichols, D. G. (2012). A qualitative study of expert and team cognition on complex patients in the pediatric intensive care unit. *Pediatric Critical Care Medicine : A Journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies, 13*(3), 278-284. doi:10.1097/PCC.0b013e31822f1766; 10.1097/PCC.0b013e31822f1766

DeShon, R. P. (2002). Generalizability theory. In F. Drasgow, & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations* (pp. 189-220). San Francisco, CA: Jossey-Bass.

Dietz, A. S., Pronovost, P. J., Mendez-Tellez, P. A., Wyskiel, R., Marsteller, J. A., Thompson, D. A., & Rosen, M. A. (2014a). A systematic review of teamwork in the intensive care unit: What do we know about teamwork, team tasks, and improvement strategies? *Journal of Critical Care, 29*(6), 908-914. doi:10.1016/j.jcrc.2014.05.025.

Dietz, A. S., Pronovost, P. J., Benson, K. N., Mendez-Tellez, P. A., Dwyer, C., Wyskiel, R., & Rosen, M. A. (2014b). A systematic review of behavioural marker systems in healthcare: What do we know about their attributes, validity and application? *BMJ Quality & Safety,* doi:bmjqs-2013-002457.

Dimick, J. B., Pronovost, P. J., Heitmiller, R. F., & Lipsett, P. A. (2001). Intensive care unit physician staffing is associated with decreased length of stay, hospital cost, and complications after esophageal resection. *Critical Care Medicine, 29*(4), 753-758.

Dodek, P. M., & Raboud, J. (2003). Explicit approach to rounds in an ICU improves communication and satisfaction of providers. *Intensive Care Medicine, 29*(9), 1584-1588. doi:10.1007/s00134-003-1815-y

Donchin, Y., Gopher, D., Olin, M., Badihi, Y., Biesky, M., Sprung, C. L., . . . Cotev, S. (1995). A look into the nature and causes of human errors in the intensive care unit. *Critical Care Medicine, 23*(2), 294-300.

Douglas, S., Cartmill, R., Brown, R., Hoonakker, P., Slagle, J., Schultz Van Roy, K., . . . Carayon, P. (2013). The work of adult and pediatric intensive care unit nurses. *Nursing Research, 62*(1), 50-58. doi:10.1097/NNR.0b013e318270714b; 10.1097/NNR.0b013e318270714b

Dwyer, D. J., Fowlkes, J. E., Oser, R. L., Salas, E., & Lane, N. E. (1997). Team performance measurement in distributed environments: The TARGETS methodology. . In M. T. Brannick, E. Salas & C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp. 137-153). Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32-64.

Figueroa, M. I., Sepanski, R., Goldberg, S. P., & Shah, S. (2013). Improving teamwork, confidence, and collaboration among members of a pediatric cardiovascular intensive care unit multidisciplinary team using simulation-based team training. *Pediatric Cardiology, 34*(3), 612-619. doi:10.1007/s00246-012-0506-2; 10.1007/s00246-012-0506-2

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*(4), 327-358. doi:10.1037/h0061470

Fletcher, G., Flin, R., McGeorge, P., Glavin, R., Maran, N., & Patey, R. (2003). Anaesthetists' non-technical skills (ANTS): Evaluation of a behavioural marker system. *British Journal of Anaesthesia, 90*(5), 580-588.

Flin, R., & Martin, L. (2001). Behavioral markers for crew resource management: A review of current practice. *The International Journal of Aviation Psychology, 11*(1), 95-118.

Flowerdew, L., Brown, R., Vincent, C., & Woloshynowych, M. (2012). Development and validation of a tool to assess emergency physicians' nontechnical skills. *Annals of*

Emergency Medicine, 59(5), 376-385.e4. doi:10.1016/j.annemergmed.2011.11.022; 10.1016/j.annemergmed.2011.11.022

Flowerdew, L., Gaunt, A., Spedding, J., Bhargava, A., Brown, R., Vincent, C., & Woloshynowych, M. (2012). A multicentre observational study to evaluate a new tool to assess emergency physicians' non-technical skills. *Emergency Medicine Journal : EMJ,* doi:10.1136/emermed-2012-201237

Fowlkes, J., Lane, N., Salas, E., Franz, T., & Oser, R. (1994). Improving the measurement of team performance: The TARGETS methodology. *Military Psychology, 6*, 47-61.

Fowlkes, J. E., & Burke, C. S. (2005). Event-based approach to training (EBAT). In N. Stanton, A. Hedge, K. Brookhuis, E. Salas & H. Hendrick (Eds.), *Handbook of human factors and ergonomics methods* (pp. 47-1-47-5). Boca Raton, FL: CRC Press.

Frengley, R. W., Weller, J. M., Torrie, J., Dzendrowskyj, P., Yee, B., Paul, A. M., . . . Henderson, K. M. (2011). The effect of a simulation-based training intervention on the performance of established critical care unit teams. *Critical Care Medicine, 39*(12), 2605-2611. doi:10.1097/CCM.0b013e3182282a98; 10.1097/CCM.0b013e3182282a98

Gaba, D. M., Howard, S. K., Flanagan, B., Smith, B. E., Fish, K. J., & Botney, R. (1998). Assessment of clinical performance during simulated crises using both technical and behavioral ratings. *Anesthesiology, 89*(1), 8-18.

Gawande, A. A., Zinner, M. J., Studdert, D. M., & Brennan, T. A. (2003). Analysis of errors reported by surgeons at three teaching hospitals. *Surgery, 133*(6), 614-621. doi:10.1067/msy.2003.169

Gordon, M., Darbyshire, D., & Baker, P. (2012). Non-technical skills training to enhance patient safety: A systematic review. *Medical Education, 46*(11), 1042-1054. doi:10.1111/j.1365-2923.2012.04343.x; 10.1111/j.1365-2923.2012.04343.x

Graham, J., Hocking, G., & Giles, E. (2010). Anaesthesia non-technical skills: Can anaesthetists be trained to reliably use this behavioural marker system in 1 day? *British Journal of Anaesthesia, 104*(4), 440-445. doi:10.1093/bja/aeq032; 10.1093/bja/aeq032

Hackman, J. R. (2002). *Leading teams: Setting the stage for great performances*. Boston, MA: Harvard Business School Publishing Corporation.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8*(1), 23-34.

Hamilton, N., Freeman, B. D., Woodhouse, J., Ridley, C., Murray, D., & Klingensmith, M. E. (2009). Team behavior during trauma resuscitation: A simulation-based performance assessment. *Journal of Graduate Medical Education, 1*(2), 253-259. doi:10.4300/JGME-D-09-00046.1; 10.4300/JGME-D-09-00046.1

Healey, A. N., Undre, S., & Vincent, C. A. (2004). Developing observational measures of performance in surgical teams. *Quality & Safety in Health Care, 13 Suppl 1*, i33-40. doi:10.1136/qhc.13.suppl_1.i33

Ho, D., Xiao, Y., Vaidya, V., & Hu, P. (2007). Communication and sense-making in intensive care: An observation study of multi-disciplinary rounds to design computerized supporting tools. *AMIA ...Annual Symposium Proceedings / AMIA Symposium.AMIA Symposium,* , 329-333.

Hull, L., Arora, S., Kassab, E., Kneebone, R., & Sevdalis, N. (2011a). Assessment of stress and teamwork in the operating room: An exploratory study. *American Journal of Surgery, 201*(1), 24-30. doi:10.1016/j.amjsurg.2010.07.039; 10.1016/j.amjsurg.2010.07.039

Hull, L., Arora, S., Kassab, E., Kneebone, R., & Sevdalis, N. (2011b). Observational teamwork assessment for surgery: Content validation and tool refinement. *Journal of the American College of Surgeons, 212*(2), 234-243.e1-5. doi:10.1016/j.jamcollsurg.2010.11.001; 10.1016/j.jamcollsurg.2010.11.001

Ilan, R., LeBaron, C. D., Christianson, M. K., Heyland, D. K., Day, A., & Cohen, M. D. (2012). Handover patterns: An observational study of critical care physicians. *BMC Health Services Research, 12*, 11-6963-12-11. doi:10.1186/1472-6963-12-11; 10.1186/1472-6963-12-11

Jain, M., Miller, L., Belt, D., King, D., & Berwick, D. M. (2006). Decline in ICU adverse events, nosocomial infections and cost through a quality improvement initiative focusing on teamwork and culture change. *Quality & Safety in Health Care, 15*(4), 235-239. doi:10.1136/qshc.2005.016576

JCR. (2004). *Improving care in the ICU*. Oakbrook Tree, IL: Joint Commission Resources (JCR).

Jeffcott, S. A., & Mackenzie, C. F. (2008). Measuring team performance in healthcare: Review of research and implications for patient safety. *Journal of Critical Care, 23*(2), 188-196. doi:10.1016/j.jcrc.2007.12.005; 10.1016/j.jcrc.2007.12.005

Johnson, V., Mangram, A., Mitchell, C., Lorenzo, M., Howard, D., & Dunn, E. (2009). Is there a benefit to multidisciplinary rounds in an open trauma intensive care unit regarding ventilator-associated pneumonia? *The American Surgeon, 75*(12), 1171-1174.

Jukkala, A. M., James, D., Autrey, P., Azuero, A., & Miltner, R. (2012). Developing a standardized tool to improve nurse communication during shift report. *Journal of Nursing Care Quality, 27*(3), 240-246. doi:10.1097/NCQ.0b013e31824ebbd7; 10.1097/NCQ.0b013e31824ebbd7

Kozlowski, S. W. J., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest, 7*(3), 77-124.

Kraiger, K., & Teachout, M. S. (1990). Generalizability theory as construct-related evidence of the validity of job performance ratings. . *Human Performance, 3*(1), 19-35.

Lamb, B. W., Sevdalis, N., Mostafid, H., Vincent, C., & Green, J. S. (2011). Quality improvement in multidisciplinary cancer teams: An investigation of teamwork and clinical decision-making and cross-validation of assessments. *Annals of Surgical Oncology, 18*(13), 3535-3543. doi:10.1245/s10434-011-1773-5; 10.1245/s10434-011-1773-5

Lamb, B. W., Wong, H. W., Vincent, C., Green, J. S., & Sevdalis, N. (2011). Teamwork and team performance in multidisciplinary cancer teams: Development and evaluation of an observational assessment tool. *BMJ Quality & Safety, 20*(10), 849-856. doi:10.1136/bmjqs.2010.048660; 10.1136/bmjqs.2010.048660

Leape, L. L., Brennan, T. A., Laird, N., Lawthers, A. G., Localio, A. R., Barnes, B. A., . . . Hiatt, H. (1991). The nature of adverse events in hospitalized patients. results of the harvard medical practice study II. *The New England Journal of Medicine, 324*(6), 377-384. doi:10.1056/NEJM199102073240605

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*, 273-307.

Lin, F., Chaboyer, W., & Wallis, M. (2009). A literature review of organisational, individual and teamwork factors contributing to the ICU discharge process. *Australian Critical Care : Official Journal of the Confederation of Australian Critical Care Nurses, 22*(1), 29-43. doi:10.1016/j.aucc.2008.11.001; 10.1016/j.aucc.2008.11.001

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*(6), 382-385.

Marks, M. A., Mathieu, J. E., & Zaccaro, J. (2001). A temporally based framework and taxonomy of team processes. *The Academy of Management Review, 26*(3), 356-376.

Mathieu, J. E., & Day, D. V. (1997). Assessing processes within and between organizational teams: A nuclear power plant example. In M. T. Brannick, E. Salas & C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp. 173-195). Mahwah, NJ: Lawrence Erlbaum Associates.

Mayer, C. M., Cluff, L., Lin, W. T., Willis, T. S., Stafford, R. E., Williams, C., . . . Amoozegar, J. B. (2011). Evaluating efforts to optimize TeamSTEPPS implementation in surgical and pediatric intensive care units. *Joint Commission Journal on Quality and Patient Safety / Joint Commission Resources, 37*(8), 365-374.

Meurling, L., Hedman, L., Sandahl, C., Fellander-Tsai, L., & Wallin, C. J. (2013). Systematic simulation-based team training in a swedish intensive care unit: A diverse response among critical care professions. *BMJ Quality & Safety,* doi:10.1136/bmjqs-2012-000994

Miller, P. A. (2001). Nurse-physician collaboration in an intensive care unit. *American Journal of Critical Care : An Official Publication, American Association of Critical-Care Nurses, 10*(5), 341-350.

Mishra, A., Catchpole, K., Dale, T., & McCulloch, P. (2008). The influence of non-technical performance on technical outcome in laparoscopic cholecystectomy. *Surgical Endoscopy, 22*(1), 68-73. doi:10.1007/s00464-007-9346-1

Mishra, A., Catchpole, K., & McCulloch, P. (2009). The oxford NOTECHS system: Reliability and validity of a tool for measuring teamwork behaviour in the operating theatre. *Quality & Safety in Health Care, 18*(2), 104-108. doi:10.1136/qshc.2007.024760; 10.1136/qshc.2007.024760

Mitchell, L., Flin, R., Yule, S., Mitchell, J., Coutts, K., & Youngson, G. (2012a). Development of a behavioural marker system for scrub practitioners' non-technical skills (SPLINTS system). *Journal of Evaluation in Clinical Practice,* doi:10.1111/j.1365-2753.2012.01825.x; 10.1111/j.1365-2753.2012.01825.x

Mitchell, L., Flin, R., Yule, S., Mitchell, J., Coutts, K., & Youngson, G. (2012b). Evaluation of the scrub practitioners' list of intraoperative non-technical skills system. *International Journal of Nursing Studies, 49*(2), 201-211. doi:10.1016/j.ijnurstu.2011.08.012; 10.1016/j.ijnurstu.2011.08.012

Moorthy, K., Munz, Y., Adams, S., Pandey, V., & Darzi, A. (2005). A human factors analysis of technical and team skills among surgical trainees during procedural simulations in a simulated operating theatre. *Annals of Surgery, 242*(5), 631-639.

Mullan, P. C., Wuestner, E., Kerr, T. D., Christopher, D. P., & Patel, B. (2012). Implementation of an in situ qualitative debriefing tool for resuscitations. *Resuscitation,* doi:10.1016/j.resuscitation.2012.12.005; 10.1016/j.resuscitation.2012.12.005

Nagpal, K., Abboudi, M., Fischler, L., Schmidt, T., Vats, A., Manchanda, C., . . . Moorthy, K. (2011). Evaluation of postoperative handover using a tool to assess information transfer and teamwork. *Annals of Surgery, 253*(4), 831-837. doi:10.1097/SLA.0b013e318211d849; 10.1097/SLA.0b013e318211d849

Narasimhan, M., Eisen, L. A., Mahoney, C. D., Acerra, F. L., & Rosen, M. J. (2006). Improving nurse-physician communication and satisfaction in the intensive care unit with a daily goals worksheet. *American Journal of Critical Care : An Official Publication, American Association of Critical-Care Nurses, 15*(2), 217-222.

Newkirk, M., Pamplin, J. C., Kuwamoto, R., Allen, D. A., & Chung, K. K. (2012). Checklists change communication about key elements of patient care. *The Journal of Trauma and Acute Care Surgery, 73*(2 Suppl 1), S75-82. doi:10.1097/TA.0b013e3182606239; 10.1097/TA.0b013e3182606239

Nishisaki, A., Nguyen, J., Colborn, S., Watson, C., Niles, D., Hales, R., . . . Nadkarni, V. M. (2011). Evaluation of multidisciplinary simulation training on clinical performance and team behavior during tracheal intubation procedures in a pediatric intensive care unit. *Pediatric Critical Care Medicine : A Journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies, 12*(4), 406-414. doi:10.1097/PCC.0b013e3181f52b2f; 10.1097/PCC.0b013e3181f52b2f

NOTSS. (2012). NOTSS: Non-technical skills for surgeons. Retrieved from http://www.abdn.ac.uk/iprc/notss

Nunnink, L., Welsh, A. M., Abbey, M., & Buschel, C. (2009). In situ simulation-based team training for post-cardiac surgical emergency chest reopen in the intensive care unit. *Anaesthesia and Intensive Care, 37*(1), 74-78.

Ohlinger, J., Brown, M. S., Laudert, S., Swanson, S., Fofah, O., & CARE Group. (2003). Development of potentially better practices for the neonatal intensive care unit as a culture of collaboration: Communication, accountability, respect, and empowerment. *Pediatrics, 111*(4 Pt 2), e471-81.

O'Leary, K. J., Boudreau, Y. N., Creden, A. J., Slade, M. E., & Williams, M. V. (2012). Assessment of teamwork during structured interdisciplinary rounds on medical units. *Journal of Hospital Medicine : An Official Publication of the Society of Hospital Medicine, 7*(9), 679-683. doi:10.1002/jhm.1970; 10.1002/jhm.1970

Osterlind, S. J. (2010). *Modern measurement: Theory principles, and applications of mental appraisal* (2nd ed.). Boston, MA: Allyn & Bacon.

OTAS. (2011). Observational teamwork assessment for surgery (OTAS). Retrieved from http://www1.imperial.ac.uk/resources/018F4A1D-5129-444E-96CF-04C524C2EA99/otas_manual.pdf

Pallant, J. (2007). *SPSS survival manual* (3rd ed.). New York, NY: McGraw-Hill.

Palma, J. P., Sharek, P. J., & Longhurst, C. A. (2011). Impact of electronic medical record integration of a handoff tool on sign-out in a newborn intensive care unit. *Journal of Perinatology : Official Journal of the California Perinatal Association, 31*(5), 311-317. doi:10.1038/jp.2010.202; 10.1038/jp.2010.202

Pascual, J. L., Holena, D. N., Vella, M. A., Palmieri, J., Sicoutris, C., Selvan, B., . . . Schwab, C. W. (2011). Short simulation training improves objective skills in established advanced practitioners managing emergencies on the ward and surgical intensive care unit. *The Journal of Trauma, 71*(2), 330-7; discussion 337-8. doi:10.1097/TA.0b013e31821f4721; 10.1097/TA.0b013e31821f4721

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, Inc.

Phipps, L. M., & Thomas, N. J. (2007). The use of a daily goals sheet to improve communication in the paediatric intensive care unit. *Intensive & Critical Care Nursing : The Official Journal of the British Association of Critical Care Nurses, 23*(5), 264-271. doi:10.1016/j.iccn.2007.02.001

Pollack, M. M., & Koch, M. A. (2003). Association of outcomes with organizational characteristics of neonatal intensive care units. *Critical Care Medicine, 31*(6), 1620-1629. doi:10.1097/01.CCM.0000063302.76602.86

Pronovost, P. J., Berenholtz, S., Dorman, T., Lipsett, P. A., Simmonds, T., & Haraden, C. (2003). Improving communication in the ICU using daily goals. *Journal of Critical Care, 18*(2), 71-75. doi:10.1053/jcrc.2003.50008

Pronovost, P. J., Thompson, D. A., Holzmueller, C. G., Lubomski, L. H., Dorman, T., Dickman, F., . . . Morlock, L. L. (2006). Toward learning from patient safety reporting systems. *Journal of Critical Care, 21*(4), 305-315. doi:10.1016/j.jcrc.2006.07.001

Reader, T. W., Flin, R., & Cuthbertson, B. H. (2007). Communication skills and error in the intensive care unit. *Current Opinion in Critical Care, 13*(6), 732-736. doi:10.1097/MCC.0b013e3282f1bb0e

Reader, T. W., Flin, R., & Cuthbertson, B. H. (2011). Team leadership in the intensive care unit: The perspective of specialists. *Critical Care Medicine, 39*(7), 1683-1691. doi:10.1097/CCM.0b013e318218a4c7; 10.1097/CCM.0b013e318218a4c7

Reader, T. W., Flin, R., Lauche, K., & Cuthbertson, B. H. (2006). Non-technical skills in the intensive care unit. *British Journal of Anaesthesia, 96*(5), 551-559. doi:10.1093/bja/ael067

Reader, T. W., Flin, R., Mearns, K., & Cuthbertson, B. H. (2009). Developing a team performance framework for the intensive care unit. *Critical Care Medicine, 37*(5), 1787-1793. doi:10.1097/CCM.0b013e31819f0451; 10.1097/CCM.0b013e31819f0451

Reader, T. W., Flin, R., Mearns, K., & Cuthbertson, B. H. (2011). Team situation awareness and the anticipation of patient progress during ICU rounds. *BMJ Quality & Safety, 20*(12), 1035-1042. doi:10.1136/bmjqs.2010.048561; 10.1136/bmjqs.2010.048561

Rehder, K. J., Uhl, T. L., Meliones, J. N., Turner, D. A., Smith, P. B., & Mistry, K. P. (2012). Targeted interventions improve shared agreement of daily goals in the pediatric intensive care unit. *Pediatric Critical Care Medicine : A Journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies, 13*(1), 6-10. doi:10.1097/PCC.0b013e3182192a6c; 10.1097/PCC.0b013e3182192a6c

Reid, J., Stone, K., Brown, J., Caglar, D., Kobayashi, A., Lewis-Newby, M., . . . Quan, L. (2012). The simulation team assessment tool (STAT): Development, reliability and validation. *Resuscitation, 83*(7), 879-886. doi:10.1016/j.resuscitation.2011.12.012; 10.1016/j.resuscitation.2011.12.012

Rosen, M. A., Schiebel, N., Salas, E., Wu, T., Silvestri, S., & King, H. B. (2012). How can team performance be measured, assessed, and diagnosed? . In E. Salas, & K. Frush (Eds.), *Improving patient safety through teamwork and team training* (pp. 59-79). Oxford, UK: Oxford University Press.

Rosen, M. A., Salas, E., Silvestri, S., Wu, T. S., & Lazzara, E. H. (2008). A measurement tool for simulation-based training in emergency medicine: The simulation module for assessment of resident targeted event responses (SMARTER) approach. *Simulation in Healthcare : Journal of the Society for Simulation in Healthcare, 3*(3), 170-179. doi:10.1097/SIH.0b013e318173038d; 10.1097/SIH.0b013e318173038d

Rosen, M. A., Salas, E., Wilson, K. A., King, H. B., Salisbury, M., Augenstein, J. S., . . . Birnbach, D. J. (2008). Measuring team performance in simulation-based training: Adopting best practices for healthcare. *Simulation in Healthcare : Journal of the Society for Simulation in Healthcare, 3*(1), 33-41. doi:10.1097/SIH.0b013e3181626276; 10.1097/SIH.0b013e3181626276

Rosen, M. A., Weaver, S. J., Lazzara, E. H., Salas, E., Wu, T., Silvestri, S., . . . King, H. B. (2010). Tools for evaluating team performance in simulation-based training. *Journal of Emergencies, Trauma, and Shock, 3*(4), 353-359. doi:10.4103/0974-2700.70746; 10.4103/0974-2700.70746

Russ, S., Hull, L., Rout, S., Vincent, C., Darzi, A., & Sevdalis, N. (2012). Observational teamwork assessment for surgery: Feasibility of clinical and nonclinical assessor calibration with short-term training. *Annals of Surgery, 255*(4), 804-809. doi:10.1097/SLA.0b013e31824a9a02; 10.1097/SLA.0b013e31824a9a02

Salas, E., Dickinson, T. L., Converse, S., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey, & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3-29). Norwood, NJ: Ablex.

Salas, E., Rosen, M. A., Curke, C. S., Nicholson, D., & Howse, W. R. (2007). Markers for enhancing team cognition in complex environments: The power of team performance diagnosis. *Aviation, Space, and Environmental Medicine, 78*(5), B77-B85.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "big five" in teamwork? *Small Group Research, 36*(5), 555-599. doi:10.1177/1046496405277134

Salas, E., Stagl, K. C., Burke, C. S., & Goodwin, G. F. (2007). Fostering team effectiveness in organizations: Toward an integrative theoretical framework of team performance. In J. W. Shuart, W. Spaulding & J. Poland (Eds.), *Modeling complex systems: Motivation, cognition and social processes. nebraska symposium on motivation* (pp. 185-243). Lincoln, NE: University of Nebraska Press.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors, 50*(3), 540-547.

Salas, E., Rosen, M. A., Burke, C. S., Nicholson, D., & Howse, W. R. (2007). Markers for enhancing team cognition in complex environments: The power of team performance diagnosis. *Aviation, Space, and Environmental Medicine, 78*(5 Suppl), B77-85.

Schmutz, J., & Manser, T. (2013). Do team processes really have an effect on clinical performance? A systematic literature review. *British Journal of Anaesthesia, 110*(4), 529-544. doi:10.1093/bja/aes513; 10.1093/bja/aes513

Sevdalis, N., Wong, H. W., Arora, S., Nagpal, K., Healey, A., Hanna, G. B., & Vincent, C. A. (2012). Quantitative analysis of intraoperative communication in open and laparoscopic surgery. *Surgical Endoscopy, 26*(10), 2931-2938. doi:10.1007/s00464-012-2287-3

Sexton, J. B., Berenholtz, S. M., Goeschel, C. A., Watson, S. R., Holzmueller, C. G., Thompson, D. A., . . . Pronovost, P. J. (2011). Assessing and improving safety climate in a large cohort of intensive care units. *Critical Care Medicine, 39*(5), 934-939. doi:10.1097/CCM.0b013e318206d26c; 10.1097/CCM.0b013e318206d26c

Shavelson, R. J., Mayberry, P. W., Li, W., & Webb, N. M. (1990). Generalizability of job performance measurements: Marine corps rifleman. *Military Psychology, 2*(3), 129-144.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Thousand Oaks, CA: SAGE Publications.

Shortell, S. M., Zimmerman, J. E., Rousseau, D. M., Gillies, R. R., Wagner, D. P., Draper, E. A., . . . Duffy, J. (1994). The performance of intensive care units: Does good management make a difference? *Medical Care, 32*(5), 508-525.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420-428. doi:10.1037/0033-2909.86.2.420

Sluiter, J. K., Bos, A. P., Tol, D., Calff, M., Krijnen, M., & Frings-Dresen, M. H. (2005). Is staff well-being and communication enhanced by multidisciplinary work shift evaluations? *Intensive Care Medicine, 31*(10), 1409-1414. doi:10.1007/s00134-005-2769-z

Smith-Jentsch, K. A., Cannon-Bowers, J. A., Tannenbaum, S. I., & Salas, E. (2008). Guided team self-correction: Impacts on team mental models, processes, and effectiveness. *Small Group Research, 39*(3), 303-327. doi:10.1177/1046496408317794

Smith-Jentsch, K. A., Zeisig, R. L., Acton, B., & McPherson, J. A. (1998). Team dimensional training: A strategy for guided team self-correction. In J. A. Cannon-Bowers, & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 271-297). Washington, DC: American Psychological Association.

Stockwell, D. C., Pollack, M. M., Turenne, W. M., & Slonim, A. D. (2005). Leadership and management training of pediatric intensivists: How do we gain our skills? *Pediatric Critical Care Medicine : A Journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies, 6*(6), 665-670.

Stockwell, D. C., Slonim, A. D., & Pollack, M. M. (2007). Physician team management affects goal achievement in the intensive care unit. *Pediatric Critical Care Medicine : A Journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies, 8*(6), 540-545. doi:10.1097/01.PCC.0000288709.15113.8A

Stone, M. E.,Jr, Snetman, D., O'Neill, A., Cucuzzo, J., Lindner, J., Ahmad, S., & Teperman, S. (2011). Daily multidisciplinary rounds to implement the ventilator bundle decreases ventilator-associated pneumonia in trauma patients: But does it affect outcome? *Surgical Infections, 12*(5), 373-378. doi:10.1089/sur.2010.067; 10.1089/sur.2010.067

Studdert, D. M., Mello, M. M., Burns, J. P., Puopolo, A. L., Galper, B. Z., Truog, R. D., & Brennan, T. A. (2003). Conflict in the care of patients with prolonged stay in the ICU:

Types, sources, and predictors. *Intensive Care Medicine, 29*(9), 1489-1497. doi:10.1007/s00134-003-1853-5

Sudikoff, S. N., Overly, F. L., & Shapiro, M. J. (2009). High-fidelity medical simulation as a technique to improve pediatric residents' emergency airway management and teamwork: A pilot study. *Pediatric Emergency Care, 25*(10), 651-656.

Sutton, G., Liao, J., Jimmieson, N. L., & Restubog, S. L. (2011). Measuring multidisciplinary team effectiveness in a ward-based healthcare setting: Development of the team functioning assessment tool. *Journal for Healthcare Quality : Official Publication of the National Association for Healthcare Quality, 33*(3), 10-23; quiz 23-4. doi:10.1111/j.1945-1474.2011.00138.x; 10.1111/j.1945-1474.2011.00138.x

Symons, N. R., Wong, H. W., Manser, T., Sevdalis, N., Vincent, C. A., & Moorthy, K. (2012). An observational study of teamwork skills in shift handover. *International Journal of Surgery (London, England), 10*(7), 355-359. doi:10.1016/j.ijsu.2012.05.010; 10.1016/j.ijsu.2012.05.010

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson Education, Inc.

Thomas, E. J. (2011). Improving teamwork in healthcare: Current approaches and the path forward. *BMJ Quality & Safety, 20*(8), 647-650. doi:10.1136/bmjqs-2011-000117; 10.1136/bmjqs-2011-000117

Thomas, E. J., Sexton, J. B., & Helmreich, R. L. (2003). Discrepant attitudes about teamwork among critical care nurses and physicians. *Critical Care Medicine, 31*(3), 956-959. doi:10.1097/01.CCM.0000056183.89175.76

Thomas, E. J., Sexton, J. B., & Helmreich, R. L. (2004). Translating teamwork behaviours from aviation to healthcare: Development of behavioural markers for neonatal resuscitation. *Quality & Safety in Health Care, 13 Suppl 1*, i57-64. doi:10.1136/qhc.13.suppl_1.i57

Tinsley, H. E. A., & Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology, 22*(4), 358-376.

Trochim, W. M. (October 20, 2006). The research methods knowledge base, 2nd edition. Retrieved from http://www.socialresearchmethods.net/kb/

Undre, S., Sevdalis, N., Healey, A. N., Darzi, A., & Vincent, C. A. (2007). Observational teamwork assessment for surgery (OTAS): Refinement and application in urological surgery. *World Journal of Surgery, 31*(7), 1373-1381. doi:10.1007/s00268-007-9053-z

Vigorito, M. C., McNicoll, L., Adams, L., & Sexton, B. (2011). Improving safety culture results in rhode island ICUs: Lessons learned from the development of action-oriented plans. *Joint Commission Journal on Quality and Patient Safety / Joint Commission Resources, 37*(11), 509-514.

Walker, S., Brett, S., McKay, A., Lambden, S., Vincent, C., & Sevdalis, N. (2011). Observational skill-based clinical assessment tool for resuscitation (OSCAR): Development and validation. *Resuscitation, 82*(7), 835-844. doi:10.1016/j.resuscitation.2011.03.009; 10.1016/j.resuscitation.2011.03.009

Weller, J., Frengley, R., Torrie, J., Shulruf, B., Jolly, B., Hopley, L., . . . Paul, A. (2011). Evaluation of an instrument to measure teamwork in multidisciplinary critical care teams. *BMJ Quality & Safety, 20*(3), 216-222. doi:10.1136/bmjqs.2010.041913; 10.1136/bmjqs.2010.041913

Westli, H. K., Johnsen, B. H., Eid, J., Rasten, I., & Brattebo, G. (2010). Teamwork skills, shared mental models, and performance in simulated trauma teams: An independent group design. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, 18*, 47-7241-18-47. doi:10.1186/1757-7241-18-47; 10.1186/1757-7241-18-47

Wetzel, C. M., Black, S. A., Hanna, G. B., Athanasiou, T., Kneebone, R. L., Nestel, D., . . . Woloshynowych, M. (2010). The effects of stress and coping on surgical performance during simulations. *Annals of Surgery, 251*(1), 171-176. doi:10.1097/SLA.0b013e3181b3b2be; 10.1097/SLA.0b013e3181b3b2be

Wetzel, C. M., George, A., Hanna, G. B., Athanasiou, T., Black, S. A., Kneebone, R. L., . . . Woloshynowych, M. (2011). Stress management training for surgeons-a randomized, controlled, intervention study. *Annals of Surgery, 253*(3), 488-494. doi:10.1097/SLA.0b013e318209a594; 10.1097/SLA.0b013e318209a594

Wilson, K. A., Salas, E., Priest, H. A., & Andrews, D. (2007). Errors in the heat of battle: Taking a closer look at shared cognition breakdowns through teamwork. *Human Factors, 49*(2), 243-256.

Yule, S., Flin, R., Paterson-Brown, S., Maran, N., & Rowley, D. (2006). Development of a rating system for surgeons' non-technical skills. *Medical Education, 40*(11), 1098-1104. doi:10.1111/j.1365-2929.2006.02610.x

Yule, S., Flin, R., Maran, N., Rowley, D., Youngson, G., & Paterson-Brown, S. (2008). Surgeons' non-technical skills in the operating room: Reliability testing of the NOTSS behavior rating system. *World Journal of Surgery, 32*(4), 548-556. doi:10.1007/s00268-007-9320-z.

Yule, S., & Paterson-Brown, S. (2012). Surgeons' non-technical skills. *The Surgical Clinics of North America, 92*(1), 37-50. doi:10.1016/j.suc.2011.11.004; 10.1016/j.suc.2011.11.004

Yule, S., Rowley, D., Flin, R., Maran, N., Youngson, G., Duncan, J., & Paterson-Brown, S. (2009). Experience matters: Comparing novice and expert ratings of non-technical skills using the NOTSS system. *ANZ Journal of Surgery, 79*(3), 154-160. doi:10.1111/j.1445-2197.2008.04833.x; 10.1111/j.1445-2197.2008.04833.x

Zaccaro, S. J., Rittman, A. L., & Marks, M. A. (2002). Team leadership. *The Leadership Quarterly, 12*, 451-483.