

AN EXPECTATION STATES APPROACH TO EXAMINING MEDICAL TEAM
INFORMATION EXCHANGE

A Thesis

by

BIANCA NICOLE MANAGO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,
Committee Members,

Head of Department,

Jane Sell
Verna Keith
Srividya Ramasubramanian
Jane Sell

August 2013

Major Subject: Sociology

Copyright 2013 Bianca Nicole Manago

ABSTRACT

This project is the first step in a long line of research that will examine the impact of status on information exchange in small groups of medical professionals. Specifically, we employ the expectation states theory and observable power and prestige methodology to develop a coding scheme and live coding methodology that is attuned to the unique status organizing process in interprofessional medical teams.

This paper begins with an explanation of the shortcomings in current research that examines medical teams. We then discuss the conceptual development of the coding scheme and methodology. Next, we establish reliability between live coders and between the transcript coders. We conclude by employing our coding scheme to examine how occupational status (physician vs. nurse) operates in medical teams, and find that our scheme possesses both criterion and face validity. Future steps include increasing our sample size to have more statistical power in detecting status differences and dropping some items from the coding scheme to increase reliability.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Jane Sell, my committee members, Dr. Verna Keith, and Dr. Srivi Ramasubramanian, and my co-investigators, Dr. Anna Johansson and Dr. Mike Howell for their guidance and support throughout the course of this research. This final product would be far less than it is without their critical questions and assessment of my work.

Thanks also to my colleagues and the department faculty and staff for supporting me during my time at Texas A&M University. Finally, I would like to thank my family and friends for their continued encouragement, support, and love.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS.....	iv
1. INTRODUCTION.....	1
2. LITERATURE REVIEW	4
2.1 What is a Medical Team?	4
2.2 Theoretical Framing for Medical Team Performance.....	5
2.3 Concepts and Definitions	7
2.4 Methodology and Measurement.....	10
3. EXPECTATION STATES THEORY	13
3.1 Observable Power and Prestige and the Burden-of-Proof Process.....	14
3.2 How Hierarchy Functions in Groups	15
4. RESEARCH QUESTION	18
5. METHODS	19
5.1 Data Access/Selection Criteria.....	19
5.2 Conceptual Development.....	21
5.2.1 Performance Outputs.....	24
5.2.2 Action Opportunities.....	28
5.2.3 Positive Evaluations.....	28
5.2.4 Negative Evaluations.....	29
5.2.5 Other	29
5.3 Coding Sequence.....	30
5.4 Reliability.....	31
5.5 Coding Methodologies.....	33
5.5.1 Transcript Coding Methodology.....	33
5.5.2 Live Coding Methodology.....	34
6. RESULTS	37
6.1 Reliability.....	37

6.2 Live vs. Transcript Methodology.....	40
6.3 Theoretical Soundness.....	40
7. CONCLUSION/DISCUSSION.....	43
ENDNOTES	47
REFERENCES.....	48
APPENDIX A	57
APPENDIX B	60

1. INTRODUCTION

To successfully provide patient care, medical professionals with different areas of expertise must work together; this process is known as “interprofessional care” (Lindgard et al., 2012). For interprofessional care to yield positive patient outcomes, teams of caregivers must work collaboratively. Thus, effective information exchange among medical team members is essential to make the decisions and take the actions necessary to save lives. Although medical team collaboration is vital for millions of patients, surprisingly little is understood about how it functions (Lingard et al., 2012). The current lack of understanding of team collaboration stems from both the absence of both an accepted, clearly defined conceptual- framework, and a reliable and valid measurement model.

This paper seeks to address these shortcomings through offering a theoretical framework by which to understand team interaction. From this theoretical framework, we develop a reliable and valid methodology that will then be used to capture information exchange in medical teams and determine the communication processes that produce the best patient care. In this way, the developed technique can provide important guidelines for which team strategies protect both the cohesion of team members and patients’ health.

Research to date has produced equivocal results regarding communication in the medical setting and the effectiveness of teamwork and training interventions. The current medical team research has reached a point where the linkages between information exchange, hierarchy, and team performance need to be clearly established. Several individual experiments demonstrate interesting findings, but the effects they display are “complex and under analyzed” (Zelditch 2007:519). By complex, we mean that several different processes

combine to produce an effect; by under analyzed, we mean that the “conceptualization of the effect has not distinguished two or more processes from each other” (Zelditch 2007:519).

Specifically, three main issues exist within the current medical team literature: (1) the definitions of key concepts lack precision, (2) the methods for measuring key teamwork constructs are neither predictive nor can they demonstrate causation, and (3) underlying status processes (i.e. occupation, gender, race) are not sufficiently addressed. Thus, to examine the specific and distinct processes that combine to impact team performance, we must embark on a research program. Based on our analysis of the literature, we believe the first step in this process is to reassess the way these concepts are both *defined* and *measured*.

To address the aforementioned shortcomings in the literature, we propose adapting theoretical frameworks and measurement tools from sociology to medical team research. Specifically, we employ the expectation states framework and associated observable power and prestige coding methodology as our foundation. For more than fifty years, sociologists (specifically social psychologists) have used this theoretical framework and associated methodology to precisely and objectively measure the ways that status hierarchy affects information exchange, and in turn, group performance (Berger and Webster 2006). Thus, the theoretical framework and methodology we recommend is derived from the expectation states tradition, but is tailored to the medical setting.

This methodology involves developing a coding scheme for analyzing interactions that is both reliable and valid (Potter and Levine-Donnerstein 1999; Babbie 2010: 530-531; Compton, Love, and Sell 2012). That is, independent researchers must be able to produce similar findings (achieve reliability) when examining information exchange in groups by adhering to the same coding scheme and methodology. Further, the coding scheme and methodology

should measure the intended phenomena and produce logically sound results, i.e. it must possess criterion validity. In summary, this study seeks to establish a valid and reliable coding scheme and associated methodology by translating sociological theories and methods to examine information exchange in interprofessional medical teams.

We begin by reviewing the current medical team literature. Second, we elucidate key social psychological concepts and address how they apply to the medical setting. We then contrast our proposed methods with current methods in the medical teamwork literature to explain why those grounded in sociological theory are particularly suited to studying interactions in interprofessional medical care teams. Next, we propose a set of methods that can be utilized to measure the power and prestige order in medical teams. Specifically, we propose applying the expectation states framework and observable power and prestige methodology to study these phenomena. Then, we explain the development of this methodology and coding scheme. We conclude by analyzing the reliability and criterion validity of the coding scheme and methodology and discuss future applications for this methodology.

2. LITERATURE REVIEW

We begin our review of the medical team performance literature by identifying the theoretical and methodological foundation for team performance research as it currently stands. Additionally, we examine the extent to which the social structure of the team is addressed in research and training. Further, we assess the current framework that informs teamwork in health care and, in this context, address the need for a different theoretical framework that offers: (1) the precise definitions of key concepts that comprise teamwork, and (2) the methods for objective measurement of these characteristics. We then utilize these methods to examine underlying processes that impact teamwork but are currently not addressed. Finally, we elucidate how the expectation states framework and methodology can be utilized to examine teamwork behaviors.

2.1 What is a Medical Team?

Within social psychology and organizational psychology, groups are distinguished from teams. In both, groups/teams must consist of two or more people, who are aware of, interdependent on, interact and share a common goal with other group or team members (Kozlowski and Ilgen 2006; Larson 2010). Beyond that, *teams* are embedded in an “encompassing organizational system, with boundaries and linkages to the broader system context and task environment” (Kozlowski and Ilgen 2006:79). That is, in the current research, teams are generally used to refer to collections of individuals in the applied setting, and groups are more often used to refer to the experimental setting where individuals’ embeddedness in a larger organizational structure is not necessarily relevant to the task at hand. Therefore, when we examine interprofessional medical teams, we are examining two

or more individuals with variable medical training, who are aware of, interdependent on, interact with and share a common goal with other medical team members. Additionally, these medical teams are embedded in a larger social and organizational structure within their hospital and medical communities at large. For the purpose of our research, it is also important to note that medical teams have members of different genders, race/ethnicities, and occupations.

2.2 Theoretical Framing for Medical Team Performance

The research on team performance is extensive and crosses many disciplines, yet many questions remain about how team-based failures or successes occur. The *functional perspective* of group processes informs the great majority of team performance research. The functional perspective comprises group process theories that seek to describe and predict group performance as a function of inputs and processes (Hollingshead et al. 2005; Larson 2010).

The exemplar for this perspective is the input-process-output (IPO) framework of team effectiveness. In this framework, *inputs* include “the characteristics of the task to be performed, the elements of the context in which work occurs, and the attitudes team members bring to a team situation” (Baker et al. 2003:10). *Process* refers to the collaboration of team/group members to achieve a specific goal or *output* (i.e. the result from teamwork). When applied to health care teams, the IPO framework provides a descriptive model of team performance, but does not have the power to predict or explain the outcome of processes under specified conditions (Turner 1991). That is because, although team processes are inherently *dynamic*, they are generally researched and explained in *static* terms (Kozlowski et al. 1999; Marks et al. 2001; Kozlowski and Ilgen 2006). As Kozlowski and

Ilgen (2006) explain “while the I-P-O model is a useful organizing heuristic, treating it as a causal model encourages taking a limited and static perspective on team effectiveness and the dynamic processes that underlie it” (p 80).

The under-analyses of the “process” component of the I-P-O model is clearly demonstrated by research conducted within the Social Decision Scheme theory. While the social decision scheme theory research program has led to important insights on the impact that task type has on the efficacy of teams performing the tasks and the optimal group size for tasks, it does not explain *how* this happens. That is, it examines the effect of independent variables (e.g. group size, task, etc) on the dependent variable (e.g. group success), but cannot explain why, how, or the way in which this happens (Larson 2011).

Another commonly employed theoretical construct in team research is the “shared mental model.” Shared mental models refer to how task-relevant information is dispersed and shared among team members (Kozlowski and Ilgen 2006). Research suggests that degree to which individuals share an idea of how a task is to be approached (i.e. shared mental model) correlates with the effectiveness of the team. While this theoretical framework is noteworthy, it fails to address how status differences among team members may impact which member’s mental model is dominant. Additionally, although this research indicates that team processes likely mitigate the relationship between shared mental models and team performance, the way these processes operate remains unknown (Minionis, Zaccaro, and Perez 1995; Mathieu et al. 2000; Kozlowski et al. 2006). This is problematic because without a clear cause/effect link, interventions cannot be developed to improve teamwork.

The functional perspective has also informed the crew resource management (CRM) based training programs used in healthcare. These training programs are adopted from

research on flight teams, but have been re-appropriated to address the function and performance of health care teams and the implications of team work on patient outcomes (Helmreich 2000; Thomas, Sexton, Helmreich 2004). A central goal of medical team research is to develop training programs to improve team functioning and in turn, patient outcomes (Awad et al. 2005). Most of the training programs integrate principles identified as requisite for effective teamwork into training strategies to improve team members' inputs, that is their knowledge, skills and attitudes (KSAs) around teamwork. These research and intervention strategies have demonstrated equivocal results, calling into question not only how studies of team performance in complex organizations are conducted, but also how teamwork and team performance are conceptualized and measured (McGrath et al. 2000; Morey et al. 2002; Kozlowski and Ilgen 2006; Neilsen et al. 2007; Larson 2011). Below, we cite specific examples of equivocal results caused by: (1) unclear concepts and definitions and (2) interpretive/descriptive (rather than quantitative/predictive) methodologies.

2.3 Concepts and Definitions

The construct "teamwork" incorporates a number of principles, which are typically framed as "coordinating mechanisms of teamwork" (Baker et al. 2003). The coordinating mechanisms generally include: team leadership, mutual performance monitoring, backup behavior, adaptability, team/collective orientation, shared mental models, synergy, mutual trust, and closed-loop communication. All of these coordinating mechanisms share common elements in their definitions. While all of these mechanisms comprise a part of what is commonly considered "teamwork," it is the subtle differences between these mechanisms that make them interesting and useful for analysis. However, the lack of precision in the use

of key concepts allows for conflicting interpretations of the same findings, which then makes it difficult to compare studies and assess the explanatory power of competing theories.

For example, Makary et al.'s (2006) analysis of nurse and physician teamwork measured *teamwork* by utilizing the *communication and collaboration* portion of the Safety Attitudes Questionnaire (SAQ) (p 746). On the questionnaire, respondents were asked to “describe the quality of *communication and collaboration*” that they experienced with different medical professionals using a Likert-type scale (p 747). Although communication and collaboration are distinct and complex concepts, they are not evaluated separately. Furthermore, communication and collaboration are conflated with teamwork making the terms synonymous with one another. Although communication and collaboration may correlate with teamwork, it is not possible to determine the relative variance that each contributes using the current methods. Since the concepts are not analytically distinct, the overlapping of terms precludes robust measurement.

The problems arising from imprecise definitions of terms are further amplified when interprofessional care teams are surveyed, and respondents are left to interpret complex concepts. In the Makary et al. (2006) study, the evaluations reflected that physicians were more satisfied than nurses with physician-nurse collaboration. When evaluated by other medical professionals, nurses received the highest overall rating of teamwork on the communication and collaboration portion of the SAQ (4.20/5) and physicians received the lowest overall rating (3.68/5). However, these findings are ambiguous due to differing perceptions of the key concept “collaboration” by nurses and physicians.

In follow-up interviews, nurses defined good collaboration as “having their opinions and input respected, whereas physicians described good collaboration as having nurses who

anticipate their needs and follow instructions” (Makary 2006:748). An individual’s societal position will likely determine how they perceive collaboration, as seen here, these differing perceptions of collaboration appear to be rooted in status and power. Thus, a measure of collaboration that allows for socially stratified individuals to interpret its’ meaning will be inexact. So, although both conceptions of “good collaboration” are valid in ordinary language, they are not precise enough for scientific inquiry. This example demonstrates how inexact language conflates several key variables, and precludes further analysis of the effect of distinct variables on team performance.

In a study examining the relationship between teamwork behaviors and patient outcomes Mazzocco et al. (2009) coded teamwork behavior using operational definitions of certain behavior domains. The domains included: briefing, information sharing, inquiry, assertion, vigilance and awareness, and contingency management (p 679). Descriptors are then provided for each domain. For example, “information sharing” is described by the following behaviors: “[i]nformation is shared; intentions are stated; mutual respect is evident; social conversations are appropriate” (Mazzocco et al. 2009:680 – see table 1). While these individual behaviors may be associated with the sharing of information, we cannot determine which of these four potential behaviors has or has not occurred by looking at the data for “information sharing.” Furthermore, both “appropriateness of conversation” and “demonstration of mutual respect” lack behavioral descriptors and are therefore hard to quantify. Thus, the lack of behavioral descriptors and imprecise definitions make an interpretation of any findings difficult.

In summary, given the significant overlap across the coordinating mechanisms, establishing reliable and valid system for methodology and measurement can be a challenge.

Precise, analytically distinct definitions are essential to robust measurement (Reader et al. 2009). By utilizing such definitions, we can control for reference dependence, i.e. the process by which behavior in a variety of settings is partially defined by a reference point that determines one's perception of the situation at hand (Schneider 2007).

2.4 Methodology and Measurement

Methods that allow respondents or independent observers to interpret and define key concepts may further contribute to measurement problems. To better understand these measurement problems, we examine research that employs respondents' self-report, respondents' report of others behaviors, and reports of independent observers.

In a study utilizing survey methods, Baggs et al. (1999) examined how medical professionals' reports of perceived collaboration correlated with patient outcomes. When controlling for disease severity, Baggs et al. (1999) found that nurses' reports of perceived collaboration were significantly correlated with patient outcomes. That is to say, when nurses reported low levels of perceived collaboration, patient outcomes were more likely to be negative (i.e. resulting in patient death or increased complications). In contrast, they found that residents' and attending physicians' reports of collaboration were not significantly associated with patient outcomes (Baggs et al. 1999). The dramatic difference of perceptions among team members suggests that perceptions and behavior may be at odds. In addition, the respondents' invested position in the group may condition their assessment of the situation.

To address the issue of subjectivity, some studies have employed researcher observation to measure communication and teamwork behaviors (Sexton, Thomas, Helmreich 2000; Thomas, Sexton, Helmreich 2004; Thomas, Sexton, Lasky, et al. 2006; Mazzocco et al. 2009). For example, Mazzocco et al. (2009) utilized

coders to assess a variety of “teamwork behaviors” in surgical teams. Coders viewed live team interactions during the different stages of surgery and determined if these behaviors associated with teamwork occurred: never, rarely, intermittently, or frequently. To measure the frequency of “teamwork behaviors” Mazzocco et al. (2009) utilize a subjective scale (e.g. frequent, intermittently, rarely, or never) that requires assessors to guess about the relevant situational context (e.g. “what is frequent?”). This method of analysis, known as latent analysis, requires more interpretation on the part of the researcher and is less desirable because this ambiguity makes it difficult to provide a meaningful interpretation of the results (e.g. manifest coding and analysis) (Potter and Levine-Donnerstein 1999; Schneider 2007; Babbie 2010: 530-531; Compton, Love, and Sell 2012;).

Like earlier observational research, Mazzocco et al. (2009) found communication to be one of the most influential determinants of patient outcomes. Specifically, Mazzocco et al. (2009) found that despite the patient’s original condition, those patients whose surgical teams exhibited less teamwork behaviors were at a higher risk for death or complications (p 682). However, due to the observational nature of this study, the authors acknowledge that their study did not establish a causal relationship (p 683).

Although one of the central goals of team research is to develop interventions, the current medical team research is not positioned to do so. To develop interventions, we need clear measures of key teamwork constructs in the medical setting. In contrast to latent analysis that requires much interpretation, we recommend the use of manifest coding and analysis to examine group interaction. Manifest coding and analysis allow researchers to quantify very precise behaviors, such as the types of statements made and by whom they are made, the length of time for which individuals speak and the number of times each type of

statement occurs and requires little or no judgment by the observer/analyst (Compton, Love, and Sell 2012). These kinds of measurements are more easily quantifiable, and potentially allow us to examine cause and effect relationships between behavior and outcomes in the future.

Baker et al. (2003) have suggested an exploration of alternative theoretical frameworks and methods to improve team process and outcomes measures. However, the intuitive appeal of a theoretical model of team performance for the medical field leaves us with a conceptual and methodological dilemma (Baker et al. 2003). This is because it is unlikely that one overarching theory can explain all of the dimensions and complexities of medical team dynamics. However, the conceptual and methodological dilemmas inherent to team performance could be addressed using a framework that explicitly addresses the social structure of the team.

We propose an integrated framework developed in sociology for measuring and analyzing the critical dimensions of team composition, process, structure, and performance whose features have high relevance to health care teams. This method treats either the individual or the team as the unit of analysis and establishes objective measures for medical team structure and decision-making. By establishing a way to compare individual characteristics with team interactions and patient outcomes, this framework enables us to design studies that identify what is truly impacting team performance.

3. EXPECTATION STATES THEORY

The method of analysis developed in the expectation states research program allows us to examine how different communication patterns impact patient outcomes. Past research suggests that status characteristics impact communication patterns and may therefore impact patient outcomes, thus making them a salient variable for analysis when examining our methods. Before we utilize these methods to examine the impact of status characteristics on group interaction patterns, we will briefly explain how status is conceived in the literature.

When individuals come together to work on any given task, certain characteristics about group members organize group interaction. Expectation states theorists have studied these processes in great detail (Berger and Fisek 1974; Berger et al. 1977; Berger et al. 1998; and Berger, Wagner and Zelditch 1985; Berger and Webster 2006). According to the theory, certain social distinctions become status characteristics when they engender performance expectations for individuals oriented towards a specific task or set of tasks.

Expectation states theory defines two categories of status characteristics; these are specific and diffuse status characteristics. Whether diffuse or specific, status characteristics have at least two differentially evaluated states that shape performance expectations for a specific task. Specific status characteristics are often reflected in the single dimension of a specific ability that people possess, for example, an extraordinary musical ability. A specific status characteristic is most salient, and most powerful, when it is directly related to the task at hand. For example, let's say a group must correctly perform cardio-pulmonary resuscitation (CPR) on an infant, and a world-renowned pediatric emergency room physician joins the team. The combination of the high specific status (world-renowned pediatric

emergency room physician) and corresponding situational context (infant CPR) causes other group members to have high performance expectations for the physician and to defer to her expertise on questions pertaining to infants and emergencies.

Diffuse status characteristics hold performance expectations of both limited and unlimited scope. In past research, diffuse status characteristics have included characteristics such as race, gender, occupation and age. Like specific status characteristics, diffuse status characteristics have at least two differentially evaluated states that shape performance expectations. Thus, the lower status individual is more likely to defer to the higher status person, and the high status individual will exhibit more power and influence within the group interactions. These types of status characteristics are considered diffuse because they are not particular to just one task, but set general expectations of performance on a variety of tasks.

3.1 Observable Power and Prestige and the Burden-of-Proof Process

Power within group interactions is generally analyzed using the power and prestige order (Berger 1992; Berger et al. 1998; Berger and Webster 2006; Berger 2007; Compton, Love, and Sell 2012). Observable power and prestige is traditionally measured through four variables including: action rates, action opportunities, influence and compliance. In an interaction, observable power and prestige can be observed when people with high status speak more often, are asked for their opinion more often, are more likely to have their input accepted by others, and when there is disagreement, are less likely to change their mind. This measure of relative influence, known as the observable power and prestige order, is a key component to understanding how status characteristics affect performance expectations (Berger 2007).

The status generalization process occurs when socially evaluated status characteristics differ among individuals and these differences lead to inferences about the specific abilities of group members. This process begins with status differences among group members that become salient. These differences result in differential expectations for individual task ability. When this occurs, the observable power and prestige order (i.e. the group member interactions) will directly reflect the differentially evaluated states of the status characteristic. That is, those who are evaluated as having less ability will also have less influence in the group interactions. This operates in the opposite direction as well (Berger et al. 1977; Webster and Foschi 1988).

The status generalization process is shown to organize interactions on the basis of status characteristics unless some other process intervenes. This is known as the ‘burden-of-proof process’ (Berger, Cohen, and Zelditch 1972). The burden-of-proof, so to speak, is the task of demonstrating that the socially evaluated status characteristic is *not* relevant to the task at hand. That is, once a group hierarchy and set of performance expectations is established, hierarchy reigns and organizes interactions until it is interrupted.

3.2 How Hierarchy Functions in Groups

Status differences bias the amount and type of information contributed by group members, thereby shaping the overall group dynamic and the resulting status hierarchy. In task-oriented groups, the nature of the task may call for greater or lesser contributions across group members creating a more or less hierarchical group structure. For simple tasks, team hierarchy could be efficient because groups may not need input from all group members to achieve success. In some instances, it may be important for a person with content expertise to be a group leader and to provide most of the input to the task.

However, complex tasks require information and expertise from a variety of individuals for successful completion. In these instances, group hierarchy may be detrimental because it limits the contributions of low status group members. Accordingly, if the task requires significant amounts of communication from all individuals, status hierarchy impedes group performance. Status hierarchy is particularly injurious when it is based on characteristics that are not task-related (i.e. gender or race). Silver, Troyer, and Cohen (2000) note in hierarchical teams:

Higher status group members are likely to generate more ideas and opinions, while lower status members are likely to avoid these forms of information and instead generate more facts and positive evaluations (i.e. low risk forms of information). *When the group's status hierarchy is not based on ability, such a bias is likely to reduce the quality of group decision-making (P. 28).*

Status hierarchy that differentially impacts performance expectations negatively impacts group decision-making because it limits the quality and quantity of information exchanged in teams, and “reduces the *diversity* of ideas, opinions, and evaluations introduced in decision making” (Silver, Troyer, and Cohen 2000:40).

The situations that the medical teams are confronted with are often complex and require the input of all provider team members. In life-saving efforts, complete information and critical thinking are essential to assessing complex situations and developing solutions. Consequently, in the medical setting vital patient information is often diffused among hierarchically ranked professionals. In these instances, the power and prestige order may prevent low status individuals from contributing to the group. Thus, if patient information is not explicitly entered into the body of knowledge the group has to work with this may pose a problem for patient safety. Therefore, it is important to understand how status and hierarchy function in groups in medical settings and if this impacts group member contributions.

Thus, the study we propose is the first step in a research agenda that examines the impact of hierarchy on group performance. Once this phenomenon is understood, we better assess the process and the characteristics of high-performing teams that are distinct from lower-performing teams.

4. RESEARCH QUESTION

Therefore, we seek to establish (1) an accepted coding scheme for coding interactions unique to the medical setting, (2) develop a coding methodology that can be used for live medical team interactions and (3) test this method by establishing criterion and face validity. This development will allow us to examine what types of interactions lead to positive and negative patient outcomes. Furthermore, once determined, we may begin to develop interventions in the form of training programs that will utilize this knowledge to improve information exchange within medical teams and advance patient care.

5. METHODS

Using taped trainings of medical professionals, we established a coding scheme that is theoretically derived from the standard measures of the power and prestige order (Bales 1950; Berger, Cohen and Zelditch 1972) but is specially tailored to an analysis of medical teams. The theoretical development and operationalization of this coding scheme took place over one year and employed the expertise of several different kinds of researchers and medical practitioners. We then developed coding schemes for coding at the same time the interaction is occurring (live coding) and coding from transcripts of interactions (transcript coding). We tested these methodologies to ensure reliability and validity. Finally, we established the criterion validity of our coding scheme by analyzing the way that an individual's occupation (i.e. nurse vs. physician) organizes the interaction of medical professionals in teams. In the methods sections below, we explain these processes in detail.

5.1 Data Access / Selection Criteria

Skill-based competence, communication and leadership strategies are established and maintained during medical trainings. Because they closely resemble a situation that might be faced in a hospital, we use videos of taped group trainings to study observable power and prestige. Specifically, we were granted IRB approval to examine taped trainings that took place at the Carl J. Shapiro Simulation and Skills Center (SASC) at Beth Israel Deaconess Medical Center (BIDMC), a teaching hospital of Harvard Medical School, between the years of 2007 and 2010. The Simulation and Skills Center offers high-fidelity medical simulation and provides training opportunities for medical professionals by replicating real-life patient

care situations that range in complexity and severity. In fact, SASC is among the most comprehensive simulation training centers in the country.

For the purposes of our study, we utilize training videos that consist of trained medical professionals (including nurses and first year residents) who are working on a simulated medical task in the SASC. Because of the exceptional quality of the high-fidelity mock operating and intensive care unit rooms, this simulated task closely resembles a situation that might be faced in the hospital. The rooms feature full-body programmable mannequins, working water and gas lines, video cameras, and observation windows with state-of-the-art teleconferencing ability. In addition, the surgical simulation area contains partial task trainers, computer-based tutorials, filmed operations, and the Virtual Patient Program. We employ videos taken during these training exercises to examine observable power and prestige.

The videos we study are of simulated training exercises in which groups are assigned a patient in the form of a full-body programmable mannequin that presents a series of symptoms that indicate a particular ailment. The videos were of different training scenarios in which the patient may display a series of differing symptoms. However, in all scenarios, the “patient” is on the verge of physiological decompensation, as indicated by a defined set of physiological criteria. The medical professionals must work as a team to determine the appropriate diagnosis and treatment for the simulated patient. Therefore, this task requires a great deal of group interaction.

The trainings we observed included two to three residents and one nurse and varied in terms of gender and race/ethnicity. There were two different training scenarios, one that included a male patient who presented symptoms associated with acute pancreatitis, and the

other was a female patient who presented with pulmonary edema. To maintain consistency, we selected the first eight to ten minutes from each training video. Once the video segments were selected, we coded the interactions for the type of statement initiated, by whom it was initiated, and for the total amount of time each individual spoke.

5.2 Conceptual Development

Bales et al. (1951) first developed a set of interrelated actions now known as the observable power and prestige order (OPPO). The original classifications were: questions, answers, and positive/negative reactions. These categories were later renamed by Berger, Cohen and Zelditch (1972) and more precisely defined by Berger and Conner (1974) to include: action opportunities, performance outputs, and negative or positive evaluations of these contributions. Action opportunities are when one person asks another (usually in the form of a question) to provide a performance output i.e. a contribution to the group's decision. These outputs are then evaluated, either positively or negatively thereby determining one's influence in the group.

These traditional measures of OPPO are essential to capturing group member influence and served as the starting point for our conceptual development. Since the development of this initial coding scheme by Bales (1950) and alterations by Berger, Cohen and Zelditch (1972), many variations have arisen which have helped inform our coding scheme (Bales 1950, 1970; Strodbeck, James, and Hawkins, 1957; Berger, Cohen and Zelditch 1972; Berger and Conner 1974; Silver, Cohen and Rainwater 1988; Cohen and Silver 1989; Silver, Cohen and Crutchfield 1994; Ridgeway, Johnson, and Diekema 1994; Silver and Troyer 1998, Silver, Troyer, and Cohen 2000; Sell, Knotterus, Mundt and Ellison

2000, Goar and Sell 2005; Ridgeway and Correll 2006; McLeer et al. 2011; Sell, Knottnerus, and Adcock-Azbill 2013).

We utilized elements from many of these past coding schemes to inform our coding scheme development. Collaborating on this scheme were: an M.D. MPH with extensive social science training, a sociology Ph.D. and expert in translational research, and a sociology graduate student. The differing educations and perspectives of the collaborators allowed for a robust scheme that was developed on theoretical principles from sociology, but suited to the hospital setting. The development of the medical-team specific coding scheme took place over the course of a little over one-year. During that time, the research team met weekly or bi-weekly and created 16 different iterations before arriving at our final scheme.ⁱ

Unlike some prior research, our coding scheme does not code for time between statements (Berger and Conner 1974; Fisek 1974; Smith-Lovin, Skvoretz, and Hudson 1986; Shelly and Webster 1997; Shelly and Munroe 1999) or gestures (Rosa and Mazur, 1979; Kollock, Blumstein, and Schwartz, 1985; Dovidio, Brown, Heltman, Ellyson, and Keating, 1988; Smith-Lovin and Brody, 1989). We omitted these types of measurements from our coding scheme to ensure reliability and theoretical soundness (McLeer et al. 2011). Additionally, we limited our analysis to statements aimed at diagnosing and treating the patient. We chose to limit our coding scheme in these ways to ensure robustness of findings. That is because the more complicated or multifaceted the coding scheme, the less reliable it may be when employed among researchers (McLeer et al. 2011; Compton, Love, and Sell 2012).

Our coding scheme was inspired by versions of the OPPO that have been applied to teams outside the experimental setting, such as the work of Silver, Troyer, and Cohen (2000). Silver, Troyer, and Cohen's (2000) research examined the impacts of status hierarchy on

information exchange in a team that was working on a company development project. In their coding scheme, Silver, Troyer, and Cohen called upon findings from their previous work (Silver, Cohen and Rainwater 1988; Cohen and Silver 1989; Silver, Cohen and Crutchfield 1994) to develop a scheme that was able to detect these status organizing processes and the effects that they had on information exchange.

“The key point is that status biases not only the amount of information a group member contributes to the group, but also the distribution of types of information across members of different statuses. Higher status group members are likely to generate more ideas and opinions, while lower status members are likely to avoid these forms of information and instead generate more facts and positive evaluations (i.e. low risk forms of information). When the group’s status hierarchy is not based on ability, such a bias is likely to reduce the quality of group decision-making” (Silver, Troyer, and Cohen 2000:27-28).

Therefore, our coding scheme and accompanying methodology is developed to detect not only the amount of information contributed by each team member, but the types of information (i.e. ideas and opinions versus facts and positive evaluations). Additionally, we pay particular attention to “risk” in our analysis. That is, “the likelihood of receiving a negative evaluation in response to initiating information in a group” (Silver, Troyer, and Cohen 2000:27). While resembling the work of Bales (1950; 1970), the work of Silver, Troyer, and Cohen (2000) differs slightly because it examines social risk (of receiving a negative evaluation) in group interaction, rather than characterizing these exchanges by social-emotional and task dimensions (Silver, Troyer, and Cohen 2000:44 see note 3).

The focus of our coding scheme is on verbal utterances (statements) that pertain directly to the task at hand. With the exception of using the direction the person is looking to interpret to whom they are speaking, we do not include physical or non-verbal gestures in our

coding scheme. Thus, in our coding scheme, we code four main types of verbal utterances, these are: action opportunities, performance outputs, and positive/negative evaluations.

Performance outputs include: directives, question directives, announcing the plan, knowledge claims, statements of fact, patient work and knowledge claim patient work. We split performance outputs into several different components to reflect the varied nature of these types of statements and how they function to impact group dynamics. Next, action opportunities include questions that solicit information (performance output) rather than an evaluation. Positive evaluations are represented by compliance, and negative evaluations are represented by disagreements. We will begin by talking briefly about the development of each type of statement within the “performance outputs” category.

5.2.1 Performance Outputs

5.2.1.1 Directives are commands given, task assignments, and task distributions that are both substantive and procedural. They are orders or directions and state “what to do” and “how to do it.” Further, directives are unsolicited but may be circumspect if they still suggest action. An example of a directive might be “Order an x-ray and an ABG” or “Let’s give her 40 of Lasix.”

Directives are a common measure of observable power and prestige in the expectation states literature, and generally indicate high status. (see Silver, Troyer, and Cohen 2000; Goar and Sell 2005). However past expectation states literature does not distinguish between different kinds of directives.

In the medical team performance literature, Hunziker et al. (2010) deal with the concept of “leadership statements” and “commands,” which include statements very similar to directives. Here, the distinction between types of directives is derived from the basis of the

content of the statement, not on how differing directives might reflect status. Thus, while our conceptualization of directives is primarily informed by the expectation states theory, we utilized knowledge from Hunziker et al. (2010) to more precisely apply this concept to the medical setting.

5.2.1.2 Question directives fall into two categories: (1) directives with an inflection that could sound like a question, or (2) questions that function to direct action rather than to solicit information. An example of a question directive might be “Will you hand me her chart?”

Originally, we did not distinguish between types of questions. However, as we continued to develop our coding scheme, we noticed a distinction between the types of question being asked. Thus, we broke the distinction into two separate categories: substantive and procedural questions. (Substantive questions were those that asked for someone’s opinion or additional information and most closely reflected the “action opportunities” category in the original measures of power and prestige. Procedural questions were those that were not soliciting information or providing an action opportunity, but were asking for someone to complete a task.)

However, this demarcation still did not seem to capture the phenomenon we were witnessing. With further analysis and insightful contributions by the physician on our team, it was revealed that many procedural questions were not questions at all, but rather commands. That is, “Can we get an x-ray?” really means, “Order an x-ray.” We call these statements *question directives*.

Question directives differ from traditional directives because question directives are a less direct way of commanding action, and thus may be more common in lower status group

members than other types of directives. Additionally, our literature review revealed past coding schemes that coded certain types of questions as performance outputs because of the responses they solicited (see Shelly's (1996) discussion of performance output tag questions and McLeer et al. (2011)).

5.2.1.3 Announcing the plan includes statements that reflect a team member's intention to follow through with the patient's plan of care. These statements demonstrate the process by which a team member asserts that they and/or their team are about to take action, or what that action is. This type of statement reflects the training that medical professionals receive and, in traditional team training terms, function as a form of closed-loop communication (Hunziker et al. 2011). An example of announcing the plan might be "I'm going to order an x-ray."

5.2.1.4 Knowledge claims are pieces of individual information (not information known, or readily available, to the rest of the group) that provide reasoning for, or evaluations of diagnoses or treatments. It can also involve a synthesis of information that results in a conclusion or diagnosis. An example might be, "Given the patient's symptoms, I think she is having an allergic reaction."

This type of statement involves a high amount of risk because the person who makes a knowledge claim is more vulnerable to negative evaluations of their performance output since it is a unique idea or opinion (Silver, Cohen and Rainwater 1988; Cohen and Silver 1989; Silver, Cohen and Crutchfield 1994; Silver and Troyer 1998; Silver, Troyer, and Cohen 2000). Risk is a concept that is commonly associated with those of high status. That is, high status individuals are more likely to take risks because they will still be positively evaluated and influential (Silver, Troyer, and Cohen 2000).

5.2.1.5 Statements of fact differ from knowledge claims because they are not a unique idea that is up for debate, but a fact, and thus, they involve less risk (Silver, Cohen and Rainwater 1988; Cohen and Silver 1989; Silver, Cohen and Crutchfield 1994; Silver and Troyer 1998, Silver, Troyer, and Cohen 2000). These include repeating information that is known or accessible to the group such as, “BP is 120/80” or “the patient’s chart says she has a history of seizures.” This type of statement is not up for debate since it is factual, thus it exposes the person who says it to less risk. Thus, we would expect these types of statements from lower status group members. In addition, statements of fact include medical specific or routine terms, for example: Tachycardic, Hypertensive, Hypotensive, Bradycardic, Tachypneic. These terms are quick ways of describing a patient’s symptoms. These are not matters of opinion nor are they diagnoses.

5.2.1.6 Patient work (other or general) includes statements to the patient, or relaying information from the patient to the rest of the team. Examples of general patient work might include “We are going to sit you up now” or “How are you feeling Ms. Harding” or “She said she’s having difficulty breathing.” Hoschild (1983) describes emotional labor as “the silent work of evoking and suppressing feeling-in ourselves and in others.” In some literature examining gendered work environments, emotional labor is discussed as something that is performed, most often by women or those of lower status, but is not recognized as “legitimate” criteria by which to be compensated or evaluated (Hochschild 1983).

5.2.1.7 Knowledge claim patient work is a knowledge claim that is expressed to the patient. This differs from traditional conceptions of patient work because it is not informing or pacifying the patient, but is rather providing a diagnosis to the patient, e.g. “We think you are having an allergic reaction.” We consider this type of statement to be very high risk as it

opens the practitioner to having their competence questioned by the patient along with other team members (Silver, Cohen and Rainwater 1988; Cohen and Silver 1989; Silver, Cohen and Crutchfield 1994; Silver and Troyer 1998; Silver, Troyer, and Cohen 2000). Therefore, we expect to find these types of statements from physicians.

5.2.2 Action Opportunities

5.2.2.1 Questions include all questions (i.e. requests for information) that do not function as a performance output (specifically question-directive) and thus reflect the “action opportunities” category of the original measures. An example of a question might be “What do you think it is?” or “Did you say 40 of Lasix?” These statements are considered low risk since they are unlikely to elicit a negative evaluation in response (Silver, Troyer, and Cohen 2000). In fact, questions seeking knowledge or advice are expected and often welcome in the medical setting (Sexton et al. 2006). Thus, we expect these types of statements from all practitioners, regardless of status.

5.2.3 Positive Evaluations

5.2.3.1 Compliance is a statement of agreement or acceptance of another’s assertion typically following a directive or knowledge claim. We coded instances of explicit compliance. That is, not every “yeah” or “okay” is considered a statement of compliance, but only those that are responding to a performance output. This is considered low risk because it is unlikely that one will receive a negative evaluation for positively evaluating another’s work. (Berger, Zelditch and Cohen, 1972; Silver, Cohen and Rainwater 1988; Cohen and Silver 1989; Silver, Cohen and Crutchfield 1994; Ridgeway, Johnson and Diekema 1994; Silver and Troyer 1998, Silver, Troyer, and Cohen 2000; Sell, Knotterus, Mundt and Ellison

2000). Lower status group members are more likely to make statements of compliance than are high status group members as they are more likely to be agreeable due to their low status.

5.2.4 Negative Evaluations

5.2.4.1 Disagreements are statements of non-compliance, disagreements, assertions that highlight the undesirable qualities of a suggestion, or rejections of another's assertion that will typically follow a directive, question or knowledge claim. We only coded very explicit examples of disagreement. For example, if a group member suggests an action "Perhaps we should give her 20 IV" and another group member responds with another action without explicitly negatively evaluating the first's idea, "yeah, or maybe start with 10 IV," we did not count this as a disagreement, but rather a performance output. A disagreement would be if they said "I don't think we should give 20, let's do 10 instead." These represent more risky statements, because they "open the door for contests between team members" possibly resulting in a negative evaluation of the person who offers the negative evaluation (Silver, Troyer, and Cohen 2000: 32). As noted by Sexton et al. (2006), in medicine "questioning someone's performance or disagreeing with their actions is taboo" (p 469). Thus, we expect that high status group members would be more likely to make these types of statements than low status group members.

5.2.5 Other

5.2.5.1 Inaudible statements were statements that could not be discerned by the live coders. The coding of these statements serves two purposes: 1) to be a surrogate for capturing speaking time even when the actual statements cannot be determined; and 2) to get a sense of what percentage of a video the live coders are able to code, and if that amount is similar for

both coders. That is, do both live coders find that about 10% of what Person 1 says cannot be coded?

To review, the following are associated with higher status: giving more performance outputs (specifically higher risk outputs e.g. knowledge claims), giving more negative evaluations (e.g. disagreements), providing fewer action opportunities (e.g. questions), and having more speaking time. The opposite is true for low status members.

Although patient work knowledge claim is association with high status, it was not clear how general patient work might be distributed within the setting. Some evidence exists that would support the idea that nurses are more likely to talk to patients than doctors because their professional expectations may require patient interaction (Sexton et al. 2006). However, other evidence suggests that patient work is gendered such that the occupation of the individual may not matter, but the gender would be an important determinant of total patient work (Hoschild 1983).

To see the coding scheme accompanied by examples in a chart, see Figure 1.

5.3 Coding Sequence

When “coding” an interaction, the live coders counted the number of times each person said each type of statement. However, if, for example, a team member made multiple directives simultaneously, e.g. “Let’s get an x-ray and an MRI.” we took the conservative approach and only counted it as one performance output (McLeer et al. 2011). Alternatively, if a person said two different types of statements in the same instance, e.g. “Start an IV. What is his BP?” we counted each type of statement (a directive and a question) separately in an effort to capture all facets of their interaction. Further, in an effort to remain conservative in

our estimates, if a person began to say a statement, was interrupted, and continued with the same type of statement, we counted this as one performance output.

5.4 Reliability

Compton, Love, and Sell (2012) describe three phases “in which reliability estimates are obtained: (1) conceptual development; (2) coder training and tentative reliability estimates, and (3) final reliability assessment” (p 5). During the first two phases, we utilized different videos than we utilized for the final study. The final reliability assessment was conducted by examining how status characteristics combine to impact communication patterns in small groups.

Two coders coded directly from transcripts (transcript coders) and two coders coded directly from video (live coders). One of the live coders also served as a transcript coder. The live coders consisted of a clinician and a non-clinician. In addition, one of the live coders is male, and both the other live/transcript coder and the transcript coder are female. The heterogeneity of coders was intentionally sought to mirror the composition of the teams and thus, prevent bias (Compton, Love, and Sell 2012). To further account for bias and possible experimenter effects, one transcript coder did not have any information about the status position of the team members. When the live coders transcribed the videos, this information was intentionally left off of the transcript for the transcript coder who was blind to status characteristics.

The aforementioned coders labeled interactions according to a specific set of measures that we refer to as the “coding scheme” (see conceptual development section and Figure 1). As mentioned above, the coders used a combination of the pretest videos and transcripts to develop this coding scheme. After many discussions, practice runs and 16

different iterations, the coders decided on a final coding scheme that was then utilized for the final reliability assessment (see Figure 1).

After the coding scheme was agreed upon, the coders practiced coding on practice transcripts. During this process, the coders established inter- and intra-coder reliability. Once reliability was achieved on the practice groups, coders began with the actual groups. Upon coding all 15 groups, we conducted a final reliability assessment for the live coders, and a separate assessment for the transcript coders. Finally, we conducted an assessment where we compared the live coding methodology to the transcript coding.

To measure reliability we used simple agreement, as recommended in Compton, Love, and Sell (2012). For assessing reliability in this circumstance, simple agreement is appropriate because for interactions lasting longer than a few minutes, the likelihood that two coders would have the same number of each statement type by chance alone is highly unlikely (Compton, Love, and Sell 2012:355). Thus, simple agreement reliability is determined by first comparing the total number of times each coder observes each type of statement for the entire data set. Then, for each type of statement, divide the total number of agreements by the largest number of observed by any given coder. The resulting value is a “relatively good indicator of reliability” when the number of possible coding categories is high (Compton, Love, and Sell 2012:355).

For example, if Coder A observed 40 directives and Coder B observed 20 directives, the agreement reliability would be $20/40 = .50$. In accordance with past research, we consider our results to be reliable if they are greater than or equal to .75, thus .50 would be considered unacceptably low (Compton, Love, and Sell 2012). See Figure 3 for coder reliability scores.

5.5 Coding Methodologies

At this point in time, transcript coding is the accepted standard methodology in OPPO research. But before we could begin the transcript coding process, the videos used for the live coding needed to be transcribed. To ensure the live-coder who also transcript coded and transcribed all of the videos had not had too much prior exposure to the videos - we did the live-coding first. However, because the live coding is the new methodology we developed, we will first discuss the traditionally accepted method (transcript coding) to provide context for the live coding methodology.

5.5.1 Transcript Coding Methodology

Transcript coding began with an extensive and thorough transcription of the training videos. Each minute of video took about ten minutes to transcribe because each minute was saturated with dialogue from many different team members. Additionally, the transcriber had to go back through and add context to the conversation, e.g. who the person was speaking to, and the intonation of each statement. Since the videos were, on average, twelve minutes, the transcriber spent about two hours transcribing each of the videos, for a total of 30 hours. Next, the transcriber removed all status characteristics in order to keep one coder blind and reduce the potential for bias. Once the transcriptions were complete, the transcript-coders began the coding process.

Training in the transcript-coding methodology and the establishment of reliability on the practice transcripts took approximately eight hours of collaborative training in addition to time spent independently coding and tallying the transcripts. This would have likely taken longer, but the transcript coders also developed the coding scheme and worked through

issues as they arose in the live coding, therefore making them intimately familiar with the coding scheme.

When coding the transcripts, coders went through each transcript line by line and wrote the statement type by each statement. For example:

P1 – Is this the EKG reading? **Q**
P3 – desat-ing a little bit **SF**
P1 - Can we get a EBG? **QD**
P4 – is there a..? (**incomplete – no code**)
P2 – You get a kit **D**
P3 – Oxygen’s still 86 **SF**

After coding the entire transcript, each coder would use the data collection sheet developed in the live coding methodology (see appendix B) to tally the codes for each medical team member. These numbers would then be used to measure reliability.

Upon establishing reliability on the practice transcripts during training, the transcript coders separately coded five transcripts and assessed reliability. At this point, the reliability was unacceptably low on many variables, so the coders went through two transcripts line-by-line to see where any discrepancies existed. This careful analysis of the transcripts led the coders to add specificity to the definitions of: directives, announcing the plan, and statements of fact (see column 3, Figure 1). The coders then went back through the original five transcripts independently and utilized the more specific coding scheme and re-assessed reliability. At this point, the reliability was acceptable, so the coders went through and independently coded the ten remaining transcripts for the final reliability assessment.

5.5.2 Live Coding Methodology

After the coding scheme was developed and agreed upon (see conceptual development, section 5.2.), we sought to establish a live coding methodology. The live coding methodology differs greatly from most coding methodologies that utilize transcripts

or tapes of conversations because it requires coders to be able to quickly identify the type of statement upon first impression. Oftentimes, the confidentiality mandated to protect patients and practitioners in the medical setting does not permit researchers to use recording devices. Thus, our hope in developing a live coding methodology is that researchers can access and analyze medical team interactions during patient care without compromising confidential patient or practitioner information. This method also saves time and money, as it does not require transcribers.

In addition to the practical advantage of live coding, Bales (1950) asserts that live observation may be superior for the observable power and prestige methodology than other methods:

"Ideally, the method is designed for use in the original observation of interaction as it occurs. There is no doubt that a certain loss of content results when the observer attempts to depend upon sound recording alone, and still another loss as the sound record is converted into written transcript. Even sound motion pictures are inferior to the original interaction[...] However, there are many sorts of problems for which an analysis of the sound recording or written transcript should prove quite adequate[...] In brief, the heart of the method is a way of classifying direct, face-to-face interaction as it takes place, act by act, and a series of ways of summarizing and analyzing the resulting data so that they yield useful information." (p. 4)

Thus, by developing a methodology that works in the live setting, we hoped to pick up on the more subtle interactions that can only be recognized by someone who is present in the room. Although the live coding methodology provides valuable information and opportunities for researchers, it can be difficult to develop a reliable live method.

To develop a coding scheme that worked for the live setting, it was necessary to practice on videos so that the coders could watch, pause, rewind, and re-watch various parts of the transcript to determine if the coding scheme was appropriate for the medical team setting. The live coders were essential to the conceptual development of the coding scheme

because they were the first to put it into practice while trying to establish reliability. Additionally, the live coders developed a data collection form (see appendix B). The development of a functional and accessible data collection form for recording live coding interactions required several iterations of the form. These variations tested different layouts, colors, and orders to determine the most efficient collection form. After the live coders developed an appropriate data collection form and established reliability during the training sessions, they were prepared for the final reliability assessment.

The live coding for the final assessment took place over a series of three sessions, which lasted for two hours each for a total of six hours. There were five groups in each session, and each group had three to four individuals and one nurse. The two live coders watched the videos simultaneously and recorded their numbers on the data collection form. These forms were then given to a third researcher who entered all data into a spreadsheet, preparing it for the statistical analysis of reliability.

6. RESULTS

The final assessment includes 15 groups for a total of 53 medical team professionals. Of these 53, 26 were female while 27 were male; 15 were nurses and 38 were physicians; and finally, 36 were white, 12 were Asian¹, and 5 were other, non-white (see Figure 2a-c). The physicians were first year medical residents, and the nurses were trained, licensed RN's. In addition, several of the nurses were also simulation center staff. Both the live and transcript coders recorded all of the statement types, i.e. directives, question directives, announcing the plan, compliance, knowledge claims, questions, statements of fact, disagreements, qualifiers, patient work knowledge claims, and patient work other.

6.1 Reliability

The first question we sought to answer is “*Have we developed a reliable coding scheme?*” Reliability was established by dividing the total number of agreed upon statements by the coder with the highest number (For this kind of coding context, this technique is considered acceptable, see Compton, Love, and Sell 2012 for discussion). As noted by Compton, Love, and Sell (2012) a score of $\geq .75$ is considered the minimum acceptable score for reliability.

We initially tried to capture varying levels of status by separating those directives that are self-directed (e.g. announcing the plan), or those that are less assertive (e.g. question directives) as being less risky, than traditional directives. However, it was more difficult to achieve reliability with these as separate constructs, so we combined directives, question directives, and announcing the plan into one category, “combined directions,” for our

analyses. This was similar in the case of patient care; patient work knowledge claim was not witnessed enough to constitute a separate analysis or to achieve high levels of reliability.

The live-coding reliability for combined directions (total .955), directives + question directives (.994), directives (.757), inaudible (.870), questions (.892), statements of fact (.783), qualifiers (.8), and combined patient work (total .981) was acceptable. However, the reliability for compliance (.682), knowledge claims (.581), and disagreements (.5) were unacceptably low (see Figure 2). This may have been because there were so few instances of these types of statements that reliability was difficult to establish. For example, between the two live-coders an average of 6 total instances of disagreement were observed.

Additionally, the role of content expertise constituted much the observed difference between the live coders. As previously noted, one coder was a physician and one was not. Due to his medical training, the physician was able to pick up on more subtle claims to knowledge that the non-physician counted as statements of fact. An example of one such subtlety was the difference between acidosis and metabolic acidosis. If the medical team member just said “acidosis” then it was just a statement of fact as they were stating the patient’s $\text{pH} < 7.35$. On the other hand, if the team member said “metabolic acidosis” it was considered a knowledge claim because application of knowledge was required to make this diagnosis. As you can see, for non-medical observers, the difference is not sharp whereas it is distinct for medical observer. For the live coding, the physician counted 62 knowledge claims and 155 statements of fact; the non-physician counted 36 knowledge claims and 198 statements of fact. When we combined the two categories, the live coders reliability was very high at .927 (Figure 3). Therefore, it appears that the difference between the coders was not

due to theoretical disagreements as to what constituted a knowledge claim or statement of fact, but rather confusion about the meaning of medical jargon.

Transcript coding reliability was established in the same manner as the live coding reliability. The reliability for combined directives (.916), directives + question directives (.952), directives (.950), question directives (.955), compliance (.949), knowledge claims (.811), questions (.952), statements of fact (.972), disagreements (.999), patient work knowledge claim (.8), patient work other (.969), and patient work total (.966) were acceptable. The transcript coders were unable to achieve reliability for announcing the plan (.677) and qualifiers (.625) (see Figure 3). We suspect that one reason why we were unable to achieve reliability on qualifiers is because there were so few.

Next, we assessed the amount of variance between the coders to ensure that the discrepancies between coders were by chance, and not due to a systematic difference between coders' perceptions of the situation. To do so, we conducted an ANOVA with the independent variable as the coder and the dependent variables were the number of times each type of statement was said. For the live and transcript coders we found low values (all > 3.585 for live, > 1.84 for transcript) of F-statistics indicating that the between-group sum of squares is small. That is to say, in the live and transcript coding "there is not much variance based on the coder, and there is more variation across observations of behaviors than there is across coders" (Compton, Love, and Sell 2012:357). Additionally, none of the findings were statistically significant ($p \leq .05$), indicating that there was a high probability that these differences were occurring by chance alone.

6.2 Live vs. Transcript Methodology

Our next question was: “*Do the two coding methodologies yield similar results?*” If it is determined that live coding obtains reliable results that vary little from transcript coding, we will have developed a new methodology for examining the same phenomenon. We calculated simple reliability for between live and transcript coding (see Figure 3). To get this value we compared the average total number of each statement type for the live coders to the transcript coders. We then took the method with the lower average and divided it by the method with the higher average for each statement type. Between the two coding schemes, only question directives (.953), total directives (.770), and knowledge claims (.980) were reliable between the two methods, the rest were not significant.

A closer examination of the data revealed that these differences were predominantly due to the fact that the transcript coding methodology was able to pick up more total interactions than the live coding methodology. Additionally, transcript coders were reliable on more statement types (9 of 10) than were the live coders (6 of 10). A better test of the compatibility of the two different methods of coding can be obtained by examining if both methods support the same substantive conclusions. We consider this question in the next section where we examine the substantive results of each method with respect to one particular characteristic of team members, occupation.

6.3 Theoretical Soundness

Upon determining reliability, our central question was: “*Does the coding scheme possess criterion validity?*” That is, to what extent are the measures patently related to concrete criteria in the “real” world? For this portion of our analysis, we examined how the independent variable “occupation” (i.e. if an individual was a physician or nurse) impacted

influence in medical teams. We chose to examine occupation rather than gender or ethnicity in this analysis because past medical research observed occupation as a significant and salient status characteristic in the medical team setting, with physicians having more influence than nurses (Makary et al. 2006, Mazzocco et al. 2009). Thus, we have data suggesting how occupational status might operate in the medical setting, but at this point, we don't have this information for other status characteristics. Of the statements for which we coded, we only analyzed those that were both reliable and statistically significant for both the live and transcript coding methodologies.

As previously noted, we expect that high status team members (physicians) will say more total directives than lower status group members (nurses). For occupation, physicians said more total directives on average than nurses (live 4.236 vs. .6 $p=.000$ and transcript 5.197 vs. 2.2 $p=.000$ – see Figure 4). Patient work also yielded very interesting results along the lines of occupation. On average, physicians engaged in more patient work when compared to nurses (live 7.066 vs. 2.633 $p=.000$ and transcript 11.35 vs. 3.833 $p=.001$). We suspect physicians may interact with patients more than nurses because physicians are expected to operate as the official representatives of the medical team.

We expected that lower status members would have, on average, more statements of fact than higher status members because they are low-risk statements that provide uncontested information to the group (Silver, Troyer, and Cohen 2006). This may be especially true for nurses as physicians often asked them to provide statements of fact by asking for updates on patient history and vital statistics. On average, nurses said three more statements of fact than physicians per group meeting (live 5.7 vs. 2.4 $p=.000$ and transcript 8.86 vs. 5.73 $p=.0034$ – see Figure 3). Thus, among nurses and physicians, status functioned

as we expected with nurses having the least number of directives and patient work, and a higher number of low risk comments such as statements of factⁱⁱ.

7. CONCLUSION/ DISCUSSION

Our analysis suggests our coding scheme and associated methodologies are able to detect status differences in medical teams with reliability. This contributes to the current literature in three important ways: (1) we achieved exceptionally high reliability on our transcript- coding methodology. Further, we believe that with more training on medical terminology (content expertise), we may be able to achieve similarly high reliability with our live-coding methodology. (2) Despite differences in reliability, it is important to note that both methodologies yielded similar results about the relationship between physicians and nurses with respect to OPP. (3) Finally, the patterns of interaction between physicians and nurses are what we would expect in the “real world” medical setting. Therefore, the coding scheme and methods can be said to possess criterion validity. Because of our robust findings, we suggest that this method can be used to evaluate what kinds of interaction lead to the best patient-care outcomes.

As noted, we have developed a coding scheme for the medical setting with high reliability. While the reliability for live coding was not as impressive as transcript coding, live coding was able to achieve acceptable reliability for those statements that are considered “high impact” or are traditionally most predictive of hierarchy in teams (e.g. directives). With the live coding, we are able to capture some subtle nuances, such as the difference between directives and question directives, which add contour to our understanding of medical team relations. However, the differences between knowledge claims and statements of fact, or those things that require content expertise are still elusive in the live coding setting.

However, this may not be necessarily true if all coders have similarly high levels of content expertise.

In the future, some of our measures may need to be more blunt and this is especially true for the live coding. For example, in future coding schemes, we may recommend that “Announcing the Plan” gets incorporated under directives or removed altogether. We might also suggest collapsing patient work and patient work knowledge claim, as there were few instances of patient work knowledge claim and again, this distinction may require extensive content expertise. It is important to note that live coding is able to achieve reliability on the high-impact statement types. Live coding is valuable because it is potentially more convenient and requires fewer resources than transcript coding (time transcribing, video or audio recording of interactions, etc.), but still yields robust findings. Furthermore, confidentiality issues are minimized in live coding as well.

Our assessment of live coding was limited because coding was not truly “live” but done by use of a video. In the future, we would like to try truly “live” coding, and/or develop better audio and video quality. Although videos provide more dimensionality than audio-recordings or transcripts, they still fall short of the genuine live setting.

Based on the information above, we are confident that the coding scheme we developed is reliable, and possesses criterion validity. Although the coding methodologies did not pick up on all of the more nuanced or less frequently occurring statement types (e.g. question directives and disagreements), the findings were profound. It is also important to note that there were 56 total observations, with only 15 nurses; therefore it is possible that with more observations, we might have more statistical power.

When discussing criterion validity, it is important to note the specific statuses associated with occupation. That is, the physicians were first year residents, rather than bona fide professional physicians with associated status and power. Also, some of the nurses were simulation staff members, thereby increasing their status in the group as they were affiliated with the simulation center and familiar with the training setting. Therefore, it is likely that the differences we observed between physicians and nurses are likely to be underestimated, as physicians will likely be of even higher status upon completing residency, and nurses who are not intimately involved with the simulation will likely have lower status.

In spite of these potential equalizers in status differences, we still find that physicians have far more observable power and prestige in the medical team setting than nurses. This adheres to expectation states theory that suggests individuals with highly evaluated diffuse status characteristics will have more observable power and prestige than their low-status counterparts. This difference in observable power and prestige is due to the performance expectations associated with the medical team members' occupational status.

Traditionally, physicians are expected to perform exceptionally well on most tasks, and especially medical tasks, because of their high levels of education and training. The designation "physician" carries a level of power and prestige that is exemplified by the type and amount of information exchanged in the group. That is, because of their confidence in their abilities, and others' generally high evaluation of their opinion, physicians were more likely to produce "high risk" statements such as knowledge claims. In fact, physicians said nearly ten times more knowledge claims on average than nurses. Additionally, physicians demonstrated their dominance in the group by stating more than twice as many directives on average than nurses.

On the other hand, the low status and concomitant performance expectations for nurses was evidenced not only by their lack of risky statements (i.e. knowledge claims) or directives, but also by the types of statements they said more frequently than physicians. For example, nurses said, on average 3 more statements of fact than physicians. These statements are relatively low risk, and thus safe for lower status members to state. Even more telling is the high amount of compliance demonstrated by nurses compared to physicians. That is, the power and prestige of physicians is not only exemplified by what they said, but also by others' reactions to their statements. Therefore, by demonstrating high levels of compliance, nurses demonstrate both their own low status, and physicians' comparatively high status and power in the medical team.

The apparent discrepancies in status between physicians and nurses are not surprising, but they may have unintended consequences. Now that we have achieved reliability and validity, we can use this method to see if more hierarchical or less hierarchical teams produce better patient outcomes. We can also see what kinds of statements, when said by whom, lead to better medical care.

This project is the first step in a research agenda targeted at improving patient outcomes by examining and modifying the information exchange process. Now that we have established a reliable and theoretically informed coding scheme that delineates the process of medical team interaction, we can begin to examine what kinds of dynamics lead to different outcomes. In some instances, it may be that more hierarchical structures lead to better outcomes; in some instances it may be that more hierarchical structures lead to worse outcomes. This coding scheme also allows us to examine whether other diffuse status characteristics such as gender and ethnicity affect team performance.

ENDNOTES

- i. Coders were all white. Of the live coders, one was male and the other was female. Both of the transcript coders were female.
- ii. While there is an overlap of gender and occupation, there is an independent and significant effect of occupation.
- iii. $\text{Total D} = \text{Directives2} + \text{Question Directives} + \text{Announcing the Plan 2}$
- iv. Some would argue that we should correct the p-values for the multiple comparisons. We could do so and even with the most conservative adjustment, the significant results remain significant.

REFERENCES

- Awad, S. S., Fagan, S. P., Bellows, C., Albo, D., Green-Rashad, B., De la Garza, M., and Berger, D. H. (2005). Bridging the communication gap in the operating room with medical team training. *American Journal of Surgery*, 190(5), 770–4.
- Baker, D., Gustafson, S., and Beaubien, J. (2003). *Medical Teamwork And Patient Safety: The Evidence-Based Link*. Rockville, Maryland: Agency for Healthcare Research and Quality.
- Bales, R. (1950). *Interaction process analysis*. Cambridge, Massachusetts: Addison Wesley.
- Bales, R., Strodtbeck, F., Mills, T., and Roseborough, M. (1951). Channels of communication in small groups. *American Sociological Review*, 461–468.
- Bales, R., and Slater, P. (1955). Role differentiation in small decision-making groups. In T. Parsons and R.F Bales (Eds.), *Family, socialization and interaction process* (pp. 259–306). Glencoe, Illinois: Free Press.
- Bales, R. (1970). *Personality and interpersonal behavior*. New York: Holt, Rinehart and Winston.
- Bell, R., Walker, H. A., and Willer, D. (2000). Power, influence, and legitimacy in organizations: Implications of three theoretical research programs. *Research in the Sociology of Organizations*, 17, 130–177.
- Berger, J., and Conner, T. (1969). Performance Expectations and behavior in small groups. *Acta Sociologica*, 12(4), 186–198.
- Berger, J., and Cohen, B. P. (1972). Status Characteristics and Social Interaction. *American Sociological Review*, 241–255.
- Berger, J., and Conner, T. (1974). Performance expectations and behavior in small groups:

A revised reformulation. In J Berger, T. Conner, and M. Fisek (Eds.), *Expectation states theory: A theoretical research program* (pp. 85–110). Cambridge, Massachusetts: Winthrop Publishers, Inc.

Berger, J. (1992). Expectations, Theory, and Group Processes. *Social Psychology Quarterly*, 55(1), 3–11.

Berger, J., Norman, R. Z., Balkwell, J. W., and Smith, R. F. (1992). Status Inconsistency in Task Situations: A Test of Four Status Processing Principles. *American Sociological Review*, 57(6), 843–855.

Berger, J., Ridgeway, C. L., Fisek, M. H., and Norman, R. Z. (1998). The Legitimation and Delegitimation of Power and Prestige Orders. *American Sociological Review*, 63(3), 379–405.

Berger, J., and M. Webster. 2006. “Expectations, Status, and Behavior.” In P. J. Burke (Ed.), *Contemporary Social Psychological Theories* (pp. 268– 300). Stanford, California: Stanford University Press.

Berger, J. (2007). The standardized experimental situation in expectation states research: Notes on history, uses, and special features. *Laboratory Experiments in the Social Sciences* (pp. 353–378).

Berwick, D. M. (2009). The epitaph of profession. *British Journal of General Practice*, 59(559), 128–131.

Coady, E. M. (1999). A strategy for nurse defibrillation in general wards. *Resuscitation*, 42(3), 183–6.

Cohen, B, and S. D. Silver. (1989). “Group Structure and information exchange: Introduction to a theory.” J Berger, M Zelditch, and B Anderson (Eds.) *Sociological Theories in Progress: New formulations* (pp. 160–181). Newbury Park, California:

Sage.

- Cohen, E. G. (1993). From theory to practice: The development of an applied research program. J. Berger and M. Zelditch (Eds.), *Theoretical Research Programs: Studies in the Growth of Theory* (pp. 385–415). Stanford, California: Stanford University Press.
- Cohen, E. G., and Lotan, R. (1995). Producing Equal-Status Interaction in the Heterogeneous Classroom. *American Educational Research Journal*, 32(1), 99–120.
- Compton, D., Love, T., and Sell, J. (2012). Developing and Assessing Intercoder Reliability in Studies of Group Interaction. *Sociological Methodology*, 42(1), 348–364.
- Cooper, S., and Wakelam, A. (1999). Leadership of resuscitation teams: "Lighthouse Leadership". *Resuscitation*, 42(1), 27–45.
- Correll, S. J., and Ridgeway, C. L. (2003). Expectation states theory. J. Delamater (Ed.), *Handbook of Social Psychology* (pp. 29–51). New York, New York: Kluwer Academic/Plenum Publishers.
- Dovidio, J., Brown, C., Heltman, K., Ellyson, S., and Keating, C. (1988). Power displays between women and men in discussions of gender-linked tasks: A multichannel study. *Journal of Personality and Social Psychology*, 4, 580–587.
- Fisek, M. H. (1974). A model for the evolution of status structures in task-oriented discussion groups. J. Berger, T. Conner, and M. Fisek (Eds.), *Expectation States Theory: A Theoretical Research Program* (pp. 53–84). Cambridge, Massachusetts: Winthrop Publishers, Inc.
- Goar, C., and Sell, J. (2005). Using task definition to modify racial inequality within task groups. *The Sociological Quarterly*, 46(815), 525–543.

- Hochschild, A. (1983). Comment on Kemper's "Social Constructionist and Positivist Approaches to the Sociology of Emotions." *American Journal of Sociology*, 89(2), 432-434.
- Hunziker, S., Bühlmann, C., Tschan, F., Balestra, G., Legeret, C., Schumacher, C., Semmer, N. K., et al. (2010). Brief leadership instructions improve cardiopulmonary resuscitation in a high-fidelity simulation: a randomized controlled trial. *Critical Care Medicine*, 38(4), 1086–91.
- Hunziker, S., Johansson, A. C., Tschan, F., Semmer, N. K., Rock, L., Howell, M. D., and Marsch, S. (2011). Teamwork and leadership in cardiopulmonary resuscitation. *Journal of the American College of Cardiology*, 57(24), 2381–8.
- Ilgen, D. R. (1999). Teams embedded in organizations: Some implications. *American Psychologist*, 54(2), 129–139.
- Institute of Medicine and the United States Committee on the Quality of Healthcare in America. (2001). *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, D.C.: National Academy Press.
- Johansson, A., and Sell, J. (2004). Sources of legitimation and their effects on group routines: A theoretical analysis. *Research in the Sociology of Organizations* (pp. 86–116).
- Johnson, C., Dowd, T. J., and Ridgeway, C. L. (2006). Legitimacy as a Social Process. *Annual Review of Sociology*, 32(1), 53–78.
- Joint Commission. (2007). Improving America's Hospitals: The Joint Commission's Annual Report on Quality and Safety. Oakbrook Terrace, Illinois. <http://jointcommission.org/>
- Kohn, L., Corrigan, J., and Donaldson, M. (2002). *To err is human: building a safer health system*. *Annales Françaises d'anesthésie et de Réanimation* (Vol. 21, pp. 453–4).

Washington, D.C.

- Kollock, P., Blumstein, P., and Schwartz, P. (1985). Sex and power in interaction: Conversational privileges and duties. *American Sociological Review*, 50(1), 34–46.
- Kozlowski, S. W. J., and Ilgen, D. R. (2006). Enhancing the Effectiveness of Work Groups and Teams. *Psychological Science in the Public Interest*, 7(3), 77–124.
- Krippendorff, K. (2004). Reliability in Content Analysis. *Human Communication Research*, 30, 411–33.
- Laughlin, P. R. (2010). Social choice theory, social decision scheme theory, and group decision-making. *Group Processes and Intergroup Relations*, 14(1), 63–79.
- Lingard, L., Vanstone, M., Durrant, M., Fleming-Carroll, B., Lowe, M., Rashotte, J., Sinclair, L. (2012). Conflicting Messages: Examining the Dynamics of Leadership on Interprofessional Teams. *Academic Medicine*, 87(12), 1–6.
- Lucas, J. (2003). Status processes and the institutionalization of women as leaders. *American Sociological Review*, 464–480.
- Makary, M. a, Sexton, J. B., Freischlag, J. a, Holzmüller, C. G., Millman, E. A., Rowen, L., and Pronovost, P. J. (2006). Operating room teamwork among physicians and nurses: teamwork in the eye of the beholder. *Journal of the American College of Surgeons*, 202(5), 746–52.
- Mannix, E. A., and Sauer, S. J. (2006). Status and power in organizational group research: Acknowledging the pervasiveness of hierarchy. *Advances in Group Processes*, 23, 149–182.
- Manser, T. (2009). Teamwork and patient safety in dynamic domains of healthcare: a review of the literature. *Acta Anaesthesiologica Scandinavica*, 53, 143–151.
- Mazzocco, K., Petitti, D. B., Fong, K. T., Bonacum, D., Brookey, J., Graham, S., Lasky, R.

- E., et al. (2009). Surgical team behaviors and patient outcomes. *American Journal Of Surgery*, 197(5), 678–85.
- McGrath, J E. (1984). Groups: Interaction and performance. Englewood Cliffs, New Jersey: Prentice-Hall.
- McLeer, J., Frederick, J., Markovsky, B., and Barnum, C. (2011). Standardizing Open Interaction Coding for Status Processes. In S. R. Thye and E. J. Lawler (Eds.), *Advances in Group Processes* (28th ed., pp. 33–58). Emerald Group Publishing Limited.
- Mohammed, S., Ferzandi, L., and Hamilton, K. (2010). Metaphor No More: A 15-Year Review of the Team Mental Model Construct. *Journal of Management*, 36(4), 876–910.
- Morey, J., Simon, R., Jay, G., Wears, R. L., Salisbury, M., Dukes, K., and Berns, S. (2002). Error Reduction and Performance Improvement in the Emergency Department through Formal Teamwork Training: Evaluation Results of the MedTeams Project. *Health Services*, 37(6), 1553–1581.
- Nembhard, I. M., and Edmondson, A. C. (2006). Making it safe: the effects of leader inclusiveness and professional status on psychological safety and improvement efforts in health care teams. *Journal of Organizational Behavior*, 27(7), 941–966.
- Ridgeway, C., and Berger, J. (1986). Expectations, Legitimation, and Dominance Behavior in Task Groups. *American Sociological Review*, 51(5), 603–617.
- Ridgeway, C. (1989). Dominance and collective hierarchy formation in male and female task groups. *American Sociological Review*, 54(1), 79–93.
- Ridgeway, C., Johnson, C., and Diekema, D. (1994). Extenal Status , Legitimacy , and Compliance in Male and Female Groups. *Social Forces*, 72(4), 1051–1077.
- Ridgeway, C., and Correll, S. J. (2006). Consensus and the creation of status beliefs. *Social Forces*, 85, 431–454.

- Ridgeway, C., and Balkwell, J. W. (2007). Group Processes and the Diffusion of Status Beliefs. *Social Psychology Quarterly*, 60(1), 14–31.
- Rodgers, D. L., Bhanji, F., and McKee, B. R. (2010). Written evaluation is not a predictor for skills performance in an Advanced Cardiovascular Life Support course. *Resuscitation*, 81(4), 453–6.
- Rosa, E., and Mazur, A. (1979). Incipient Status in small groups. *Social Forces*, 58, 18–37.
- Sell, J., Ellison, C., Knottnerus, J., and Mundt, H. (2000). Reproducing social structure in task groups: The role of structural ritualization. *Social Forces*, 79(2), 453–475.
- Sell, J., Lovaglia, M. J., Mannix, E. a., Samuelson, C. D., and Wilson, R. K. (2004). Investigating Conflict, Power, and Status Within and Among Groups. *Small Group Research*, 35(1), 44–72.
- Sexton, J., Holzmueeller, C., and Pronovost, P. (2006). Variation in caregiver perceptions of teamwork climate in labor and delivery units. *Journal of Perinatology*, 463–470.
- Shelly, R., and Munroe, P. (1999). Do women engage in less task behavior than men? *Sociological Perspectives*, 42(1), 49–67.
- Shelly, R., and Webster, M. (1997). How formal status, liking, and ability status structure interaction: Three theoretical principles and a test. *Sociological Perspectives*, 40(1), 81–107.
- Silver, S. D., Cohen, B. P., and Crutchfield, J. H. (1994). Status differentiation and information exchange in face-to-face and computer-mediated idea generation. *Social Psychology Quarterly*, 108–123.
- Silver, S. D., Troyer, L., Silver S. D., & Cohen B. P. (2000). Effects of status on the exchange of information in team decision-making: when team building isn't enough. *Advances in Interdisciplinary Studies of Work Teams*, 7(2), 30.

- Silver, S. D., Cohen, B., and Rainwater, J. (1988). "Group structure and information exchange in innovative problem solving." *Advances in Group Processes*, 5, 169– 194.
- Singh, H., and Thomas, E. (2007). Medical Errors Involving Trainees: A Study of Closed Malpractice Claims from Five Insurers. *Archives of Internal Medicine*, 167(19), 2030–2036.
- Smith-Lovin, L., Skvoretz, J. V, and Hudson, C. G. (1986). Status and Participation Groups in Six Person Groups: A Test of Skvoretz's Comparative Status Model. *Social Forces*, 64(4), 992–1005.
- Smith-Lovin, L., and Brody, C. (1989). Interruptions in group discussions: The effects of gender and group composition. *American Sociological Review*, 54(3), 424–435.
- Thomas, E. J. (2004). Translating teamwork behaviours from aviation to healthcare: development of behavioural markers for neonatal resuscitation. *Quality and Safety in Health Care*, 13, i57–i64.
- Thomas, E. J., Taggart, B., Crandell, S., Lasky, R. E., Williams, a L., Love, L. J., Sexton, J. B., et al. (2007). Teaching teamwork during the Neonatal Resuscitation Program: a randomized trial. *Journal of Perinatology: Official Journal of the California Perinatal Association*, 27(7), 409–14.
- Troyer, L. (2004). Democracy in a Bureaucracy: The legitimacy paradox of teamwork in organizations. *Research in the Sociology of Organizations*, 22, 49–87.
- Tschan, F., Semmer, N. K., Gurtner, A., Bizzari, L., Spychiger, M., Breuer, M., and Marsch, S.U. (2009). Explicit Reasoning, Confirmation Bias, and Illusory Transactive Memory:

A Simulation Study of Group Medical Decision Making. *Small Group Research*, 40(3), 271–300.

Webster, M., and Sell, J. (2011). Groups and Institutions, Structures and Processes. In G. Ritzer (Ed.), *The Wiley-Blackwell Companion to Sociology* (p. 139). Chichester, UK: Wiley- Blackwell.

(<http://onlinelibrary.wiley.com/doi/10.1002/9781444347388.ch8/summary>).

Zelditch, M. J. (2007). Laboratory Experiments in Sociology. M. J. Webster and J. Sell (Eds.), *Laboratory Experiments in the Social Sciences* (1st ed., pp. 517–530). Oxford, England: Elsevier.

Zeltser, M. V, and Nash, D. B. (2010). Approaching the evidence basis for aviation-derived teamwork training in medicine. *American Journal of Medical Quality*, 25(1), 13–23.

APPENDIX A

Coding Scheme		
Statement Type	Definition	Transcript Clarification
Directive	Directives are both substantive and procedural. They are orders or directions and state “what to do” and “how to do it.” They include commands given, assignment of tasks, and task distribution. Example – <i>Let’s order an x-ray and an ABG</i>	(1) Directives are unsolicited assertions. Example – <i>Should I start the IV?</i> <i>Yes.</i> – NOT a directive. (2) Directives may be circumspect if they are prospective and suggest action Example - <i>So, we can give her 40mls of ...</i> (D)
Question Directives	Expresses request in the form of a question. Directives with an inflection that sounds like a question will be coded as directives. Example – <i>Can you hand me her chart?</i>	
Announcing the Plan	Asserting that the person is about to take action, or what that action is Example (spoken to other team members) – <i>I’m going to order an x-ray</i>	
Compliance	Statements of agreement typically following a directive or knowledge claim. Example – <i>Yes. I agree.</i> etc. Fillers will not be coded.	Includes responses to questions that are seeking confirmation of an action to take. Example: <i>Can we get an EKG? (QD)</i> <i>Yes. (C)</i>
Knowledge Claims	Individual information (not information known, or readily available, to rest of the group) that provides reasoning for or evaluations of diagnoses or treatments. Example – <i>Given the blood pressure and heart rate, I think it is ...</i>	
Questions	All questions that do not function as a directive. Example – <i>What do you think? Does this look like pulmonary edema?</i>	These include repeating back directions from another team member. Example - <i>Start 40 of Lasix. (D)</i> <i>40? (Q)</i>
Statements of Fact	Information known to the group. Information read from chart of telemetry. Example – <i>The blood pressure is 120/80 and respiration rate is 10...</i>	Includes responses to questions that are seeking information. Example: <i>How is her urine output? It’s not coming up. (SF)</i> Example: <i>Did she get all her meds?</i> <i>Yes. (SF)</i> Does NOT include saying that something is present. Example: “Here is your 20 of Lasix, okay?” – NOT a SF – no code Does NOT include informing the group of an action recently done Example: “I just called her son.” – NOT a SF – no code
Disagreements	Statements of non-compliance, disagreements, assertions that highlight the undesirable qualities of a suggestion, or rejections of another’s assertion that typically follows a directive, question or knowledge claim. Example – <i>I don’t think we need a chest x-ray</i>	
Qualifiers	Statements that precede or follow and modify a previous statement Example – <i>It looks like – A-Fib to me. But, I don’t really know.</i> <i>OR I’m not really sure, but it looks like...</i>	
Inaudible	During the live coding, we code each time a statement is made that is inaudible. This will allow us to analyze what portion of the conversation we were unable to analyze.	

...Coding Scheme Cont'd		
Statement Type	Definition	Transcript Clarification
Patient Work (Other)	Statements to the patient, or relaying information from the patient to the rest of the team. Example – <i>We are going to sit you up now and give you oxygen.</i>	
Knowledge Claim Patient Work	A knowledge claim that is expressed to the patient. Example – <i>I think you are having an allergic reaction.</i>	
Figure 1.		

Medical Team Composition Occupation - Gender		
	Physician	Nurse
Male	19	8
Female	19	7
Figure 2a.		

Medical Team Composition Race/Ethnicity - Gender			
	White	Asian	Other
Male	18	5	4
Female	18	7	1
Figure 2b.			

Medical Team Composition Race/Ethnicity - Occupation			
	White	Asian	Other
Physician	21	12	5
Nurse	15	0	0
Figure 2c.			

Statement Type	Reliability Assessments of Coders and Coding Methodologies		
	Live Coding	Transcript Coding	Live vs. Transcript ¹
Directives	.7570**	.9504**	0.684
Question Directives	.6621	.9545**	0.953**
Announcing the Plan	.7059	.6774	0.547
Combined Directions	.9767**	.9163**	0.742*
Directives + Question Directives	.9936**	.9517**	0.770**
Compliance	.6818	.9495**	0.192
Knowledge Claims	.5806	.8113**	0.980**
Question	.8923**	.9522**	0.430
Statement of Fact	.7828**	.9721**	0.430
Disagreement	.5	.9999**	0.333
Qualifiers	.8*	.625	0.692
Patient Work Knowledge Claim	.4286	.8**	0.6
Patient Work Other	.9403**	.9691**	0.612
Patient Work Total	.9807**	.9657**	0.632
¹ Live vs. Transcript reliability was calculated by dividing the low by high mean number of each statement type in each methodology (Compton, Love, and Sell 2012). Ex. Average of 170 total directives in live-coding [(168 + 172)/2] and 229 [(239 + 219)/2] in transcript-coding, 170/229 = .742.			
Figure 3.			

Comparison of Means Tests						
Statement Type	Live			Transcript		
	Nurse	Physician	P Value	Nurse	Physician	P Value
Total Directives ⁱⁱⁱ	.6	4.236	.0000**	2.2	5.197	.0027**
Directives + Question Directives	.3667	3.947	.0000**	.9	5	.0000**
Compliance	.6333	.2368	.0247**	2.53	1.55	.0648*
Knowledge Claims	.1333	1.2368	.0014**	.0667	1.263	.0018**
Question	1.733	2.5526	.0869*	5.3	5.434	.8720
Statement of Fact	5.667	2.408	.0000**	8.866	5.73	.0034**
Disagreement	.066	.1316	.4376	.066	.066	.9919
Patient Work Total	2.6333	7.066	.0006**	3.833	11.355	.0008** ^{iv}
Total Number: 15 Nurses – 38 physicians						
Figure 4.						

APPENDIX B

Group #:

Inaudible I					I
	PERSON 1	PERSON 2	PERSON 3	PERSON 4	
Code					
Directives D					Directives
Question-Directive QD					
Announcing Plan A					A
Compliance C					C
Knowledge Claim KC					KC
Question Q					Question/Answer
Statement of Fact SF					
Disagreement Dis					Miscellaneous
Qualifier Qa					
Patient Work – Knowledge Claim PWKC					Patient Work
Patient Work – Other PWO					