REAL EVENT LEARNING AND ANALYSIS (REAL):

ASSESSING AND IMPROVING SURGICAL TEAM PERFORMANCE

PETER M. WATERS MD

A Thesis Submitted to the Faculty of

The Harvard Medical School

in Partial Fulfillment of the Requirements

for the Degree of Master of Medical Sciences in Medical Education

Harvard University

Boston, Massachusetts.

May, 2019

Thesis Mentors: Peter Weinstock MD PhD,

Christopher Roussin PhD Author: Peter M. Waters MD

Real Event Learning and Analysis (REAL):

Assessing and Improving Surgical Team Performance

Abstract

Errors in surgery can have serious consequences to the patient's health, surgical teams' liability, and hospital reputation. Despite Joint Commission on Accreditation of Healthcare Organizations (JACHO) clear mandates and regulations to lessen risk, errors continue to occur at significant rates for unclear reasons. A stark example, wrong site/wrong procedure/wrong patient (WSPEs), are incidents that should be preventable or at least made exceedingly rare. US Department of Health and Human Resources Agency for Healthcare Research and Quality (AHRQ) maintains a WSPEs should constitute a "never" event. However, WSPEs continue to occur despite World Health Organization (WHO) structured universal protocols designed to prevent them.

By establishing highly reliable cultures of individual and team training, simulation can be effective in improving communication, leadership, task management, and situational awareness. An evolution of this process is to leverage these psychological safety strategies and debriefing skills to understand live performance at the point of clinical care delivery. (Figure 1)

The REAL (Real Event Analysis and Learning) live observational and audiovisual analysis project was developed to achieve this goal - namely to create a transformable system to

achieve strong proof of concept, study and evaluate the benefits of live video capture in the operating rooms as an important patient safety/QI/QA target. REAL required many complex, iterative steps to implementation including: (1) approval and support by hospital executive, anesthesia, surgical, nursing and legal leadership; (2) hospital IRB classification and approval of REAL as a quality improvement project; (3) acceptance and active participation by the operating room professional staff in REAL; and (4) technical expertise to record 360 degree visual and auditory activities from start to finish of each procedure observed and analyzed. All were achieved.

Our findings in the Phase I pilot (3 surgical days and 11 operative cases by single surgeon) confirmed feasibility of REAL. Our findings in Phase II involved 8 operative days ((> 50 hours) of skin to skin surgery time) of surgical care for 24 patients by 4 different surgeons and multiple, variable anesthesia and nursing teams demonstrated: (1) live OR analysis is feasible and scalable; (2) team performance as assessed live in OR and by audiovisual analysis is comparable; (3) staff acceptance of live performance analysis and learning occurred, including dedication to betterment as evidenced in debriefings; (4) performance of universal protocols is variable and often incomplete; and (5) non-technical performance by nursing, anesthesia, and nursing was high but there were both at risk (safety 1) and commendable unexpected (safety 2) behaviors. Thus, audio/visual capture of live events even in an institution dedicated to patient safety and high reliability and performing at a high level as noted by validated tools, there is still risk of WSPEs, and less serious errors. Next steps are to build on our now higher-level platform of psychology safety in REAL to achieve growth in operating room team learning and performance, to further lessen error risk, enhance safety, and build even more reliable care teams.

Table of Contents

Figures

Figure 1: Schematic of Real Event Learning Iteratively with Simulation Team Training

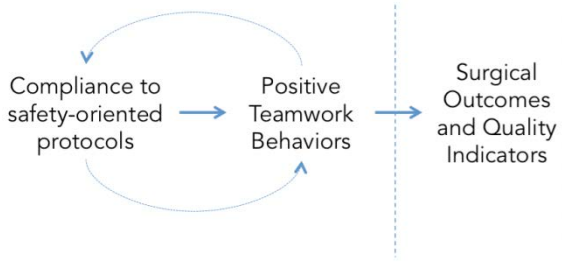


Figure 2

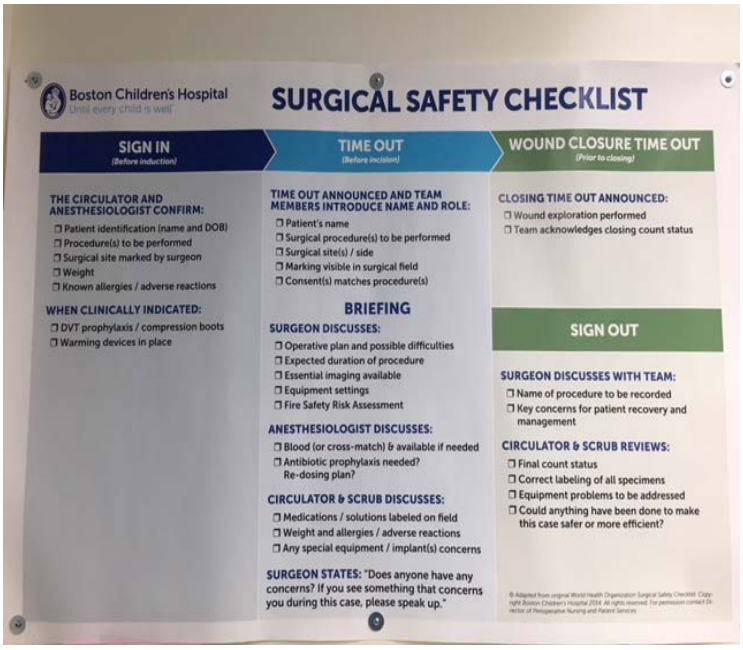


Figure 3: Visual of 4 camera view of REAL and Scoring Tools Used

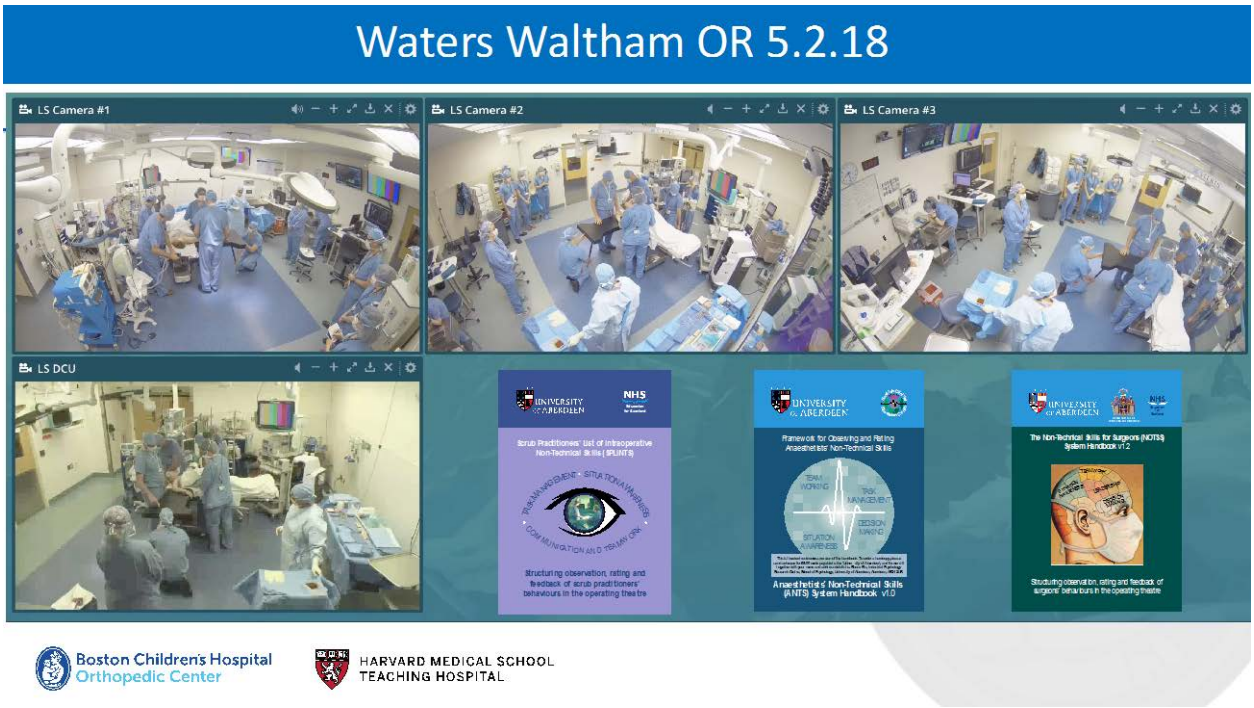
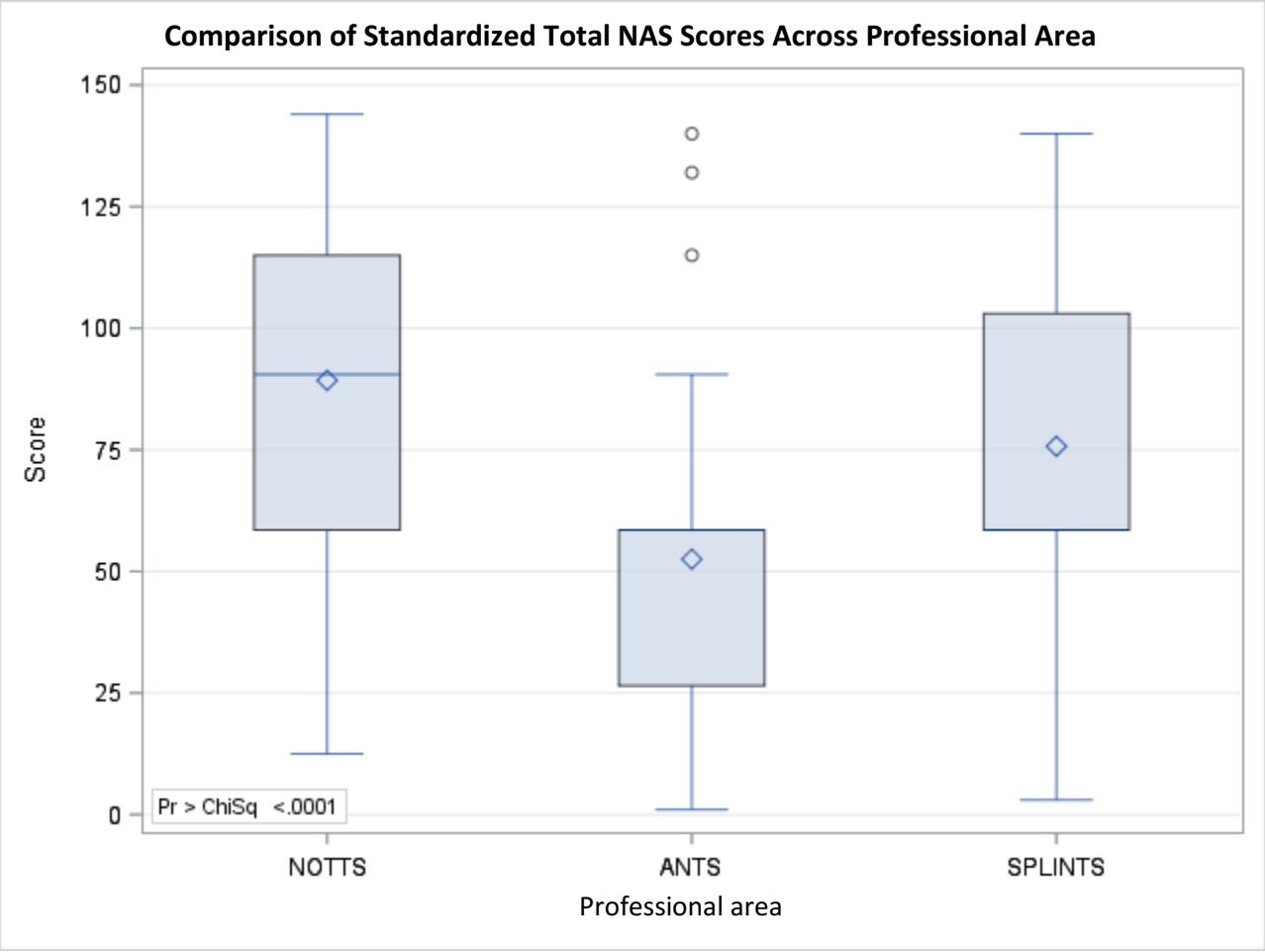


Figure 4: Comparison of Scoring by team (Prof 1 = surgeon, Prof 2 = anesthetist, Prof 3 = nurse) for Validated NOTSS, ANTS, SPLINTS (Adjusted NAS scores to account for 4 categories for NOTSS and ANTS and 3 categories for SPLINTS)



Tables

Variable	n	Missing	Total	Mean	Min	Max	Median	25th Pctl	75th Pctl
Surgeon situational awareness	24	0	80.0	3.3	2.0	4.0	3.0	3.0	4.0
Surgeon task management	0	24	.	.	.	.	.	.	.
Surgeon team skills	24	0	86.0	3.6	3.0	4.0	4.0	3.0	4.0
Surgeon decision making	24	0	90.0	3.8	3.0	4.0	4.0	3.5	4.0
Surgeon leadership	24	0	85.0	3.5	3.0	4.0	4.0	3.0	4.0
Anesthetist situational awareness	24	0	77.0	3.2	2.0	4.0	3.0	3.0	4.0
Anesthetist task management	24	0	72.0	3.0	2.0	4.0	3.0	3.0	3.0
Anesthetist team skills	24	0	67.0	2.8	2.0	3.0	3.0	3.0	3.0
Anesthetist decision making	24	0	74.0	3.1	3.0	4.0	3.0	3.0	3.0
Anesthetist leadership	0	24	.	.	.	.	.	.	.
Nurse situational awareness	24	0	78.0	3.3	3.0	4.0	3.0	3.0	3.5
Nurse task management	24	0	82.0	3.4	2.0	4.0	3.0	3.0	4.0
Nurse team skills	24	0	72.0	3.0	2.0	4.0	3.0	3.0	3.0
Nurse decision making	0	24	.	.	.	.	.	.	.
Nurse leadership	0	24	.	.	.	.	.	.	.

Table 1 NOTTS, ANTS, SPLINTS Validated (non-modified) Adjusted Video Scores

TIMEOUT Marking Visible	**24 (100%)**
TIMEOUT Team Introductions	**19 (79.2%**
TIME OUT Procedure Performed	**16 (66.67%)**
TIME OUT Surgical Site ID	**22 (91.67%)**
TIME OUT Consent Match	**24 (100%)**
BRIEFING SURGEON Operative Plan	**22 (91.67%)**
BRIEFING SURGEON Expected Time	**17 (70.8%)**
BRIEFING SURGEON Equipment Needed	**21 (87.5%)**
Briefing Anesthetist Antibiotics	**19 (79.2%**
Team Allergies	**20 (83.3%)**
Team Safety Concerns	**21 (87.5%)**
<u>While the lowest frequency time-outs were:</u>	
Team Speak-up	**10 (41.67%)**
BRIEFING SURGEON Fire Risk	**10 (41.67%)**
BRIEFING SURGEON Imaging Reviewed	**10 (41.67%)**
BRIEFING SURGEON Possible Difficulties	**8 (33.33%)**

Table 2: Compliance Scoring: Phase II 24 operations, 4 surgeons (2 days each), 8 surgical days, >50 hours surgical observation from "wheels in to wheels out" of patient and professionals in the room for each case.

NAS score	Validated (non modified) (median, q1, Q3)	Non-validated (Modified) (median, q1, Q3)	P value
Total NAS	3.27 (3.10,3.33)	3.23 (3.1,3.36)	0.273
Total NOTTS	3.6 (3.2,3.8)	3.5 (3.25, 3.75)	0.664
Total ANTS	3.0 (3.0,3.2)	3.0 (2.75,3.12)	0.917
Total SPLINTS	3.2 (3.0,3.3)	3.33(3.0,3.33)	0.666

Table 3: Comparison of Validated NOTSS, ANTS, SPLINTS (11 categories) to Non-Validated NOTSS, ANTS, SPLINTS (15 categories)

NAS non-validated score	NOTTS (median, q1, Q3)	ANTS (median, q1, Q3)	SPLINTS (median, q1, Q3)	P value
Total Score	18 (16,19)	15 (15,16)	16 (15,16.5)	<.0001
Situational awareness	3 (3,4)	3 (3,4)	3(3,3.5)	0.671
Task management	3 (3,4)	3 (3,3)	3 (3,4)	0.008
Decision making	4.0 (3.5,4)	3(3,3)	3 (3,3)	<.0001
Team skills	4 (3,4)	3 (3,3)	3 (3,3)	<.0001
Leadership	4 (3,4)	3 (3,3)	3 (3,3)	0.001

Table 4: Comparisons of NOTSS, ANTS, SPLINTS modified, non-validated,

scores

Modified (Non-validated)			
variable	Video (median, q1, Q3)	In-OR (median, q1, Q3)	P value
Total NAS	49.0 (46.5,50.0)	44.5 (39.5, 50.0)	0.008
Situational awareness	9.5 (9.0,10.0)	9.0 (9.0,10.0)	0.258
Task management	10 (9.0,10.0)	9.0(7.0,10.0)	0.026
Leadership	10 (9.0,10,0)	9.0 (8.0,10.0)	0.051
Decision making	10.0 (9.5,10.0)	9.0 (7.5, 10.0)	0.012
Teamwork	9.5 (9.0,10.0)	9.0 (8.0,10.5)	0.917
Non-modified (Validated)			
Total NAS	36 (34.0,37.0)	34.0 (32.0, 36.5)	0.046
Situational awareness	9.5 (9.0,10.0)	9.0 (9.0,10.0)	0.322
Task management	6 (6.0,7.0)	6.0 (6.0,7.0)	0.383
Leadership	4.0 (3.0, 4.0)	3.0 (3.0,4.0)	0.091
Decision making	7.0 (7.0,7.0)	6.0 (6.0,7.0)	0.012
Teamwork	9.5 (8.0,10.0)	9.0 (8.0,10.5)	0.932

Table 5: Comparison of Video Remote Scoring and in-OR Scoring for both Validated (11 categories) and Non- Validated (15 categories) NOTSS, ANTS, SPLINTS Standardized scores

Acknowledgements

I am blessed to work with and be inspired by many highly trained, inquisitive professionals. My fellow HMS MMSc-Medical Education students, especially our "Blonde Ambition" team, were very helpful in critiquing and testing the methods in the formative stages of this project. My group project team at the HMS MACY Institute led by Holly Gooding MD further "kicked the tires" of this project and made it better. My thesis committee of James A. Gordon MD, Steven Yule PhD, and Dr. J. Lawrence Marsh are world experts on simulation performance learning and proved invaluable with constructive review and guidance throughout the process. They not only are esteemed professional colleagues and mentors, but they have become friends along the way. The deep commitment of my thesis mentors Peter Weinstock MD PhD and Christopher Roussin PhD, along with Catherine Allan MD of the Boston Childrens SimPEDS program, to build and nourish a platform of psychology safety in simulation led us to the Real Event Analysis and Learning (REAL) work in our orthopaedic and sports medicine operating rooms in Waltham that provided the data for my thesis. With the entire SimPEDs team, we iteratively advanced our thoughts, tested them in group think, analyzed the outcome, and improved our product at each phase. The SimPEDS technical team of Kaitlyn Nogueria, Kelsey Graber, Duncan Smith-Freedman, and Megan Garafalo did yeoman's audiovisual recording and Learning Space analytic storage space work. Hours and hours and hours were spent by them unobtrusively watching, listening, and capturing live surgical teams at work. They also were valuable members of the strategic planning team. Gaia Uman joined as time went by and was essential in data organization and analysis. Laura Lins and Jennifer Kallini "volunteered" a significant aspect of their medical school research year on this project as coders of the audiovisual tapes of the anesthesia, surgical, and nursing operating room teams. Their input made our analysis possible; their thoughtful recommendations made the process better and more

sustainable. Along the way, I think they learned how to be better surgeons and team leaders. Ben Shore MD MPH was "coerced" to see this a next step on his career path as a leader in surgical education and eventually faculty development. It cost him valuable time but I hope opened his eyes to the possibilities. He is a true partner of mine. David Williams PhD came on board in Phase II, worked on the quantitative analysis, and became a partner in our debriefing qualitative analysis. I believe in teams, have my entire professional life, and this research team of ours did outstanding, groundbreaking work in my opinion. The future is bright and I believe this is much more to come beyond these pages.

Real event analysis and learning is not possible without institutional support at all levels. At the highest level of Boston Children's leadership, Sandi Fenwick, Kevin Churchwell, Michele Garvin, Jim Kasser, Laura Wood, and Paul Hickey endorsed and supported this project from the outset. They continue to lead with vision and courage. CRICO endorsed the project and provided the necessary financial support with a two-year grant. The Boston Children's Institutional Review Board and Legal teams met with us often to be certain we protected our patients, families, staff, and the enterprise as a whole. Their open minds and professional commitment to safety paved a healthy path of quality improvement for us. The professionals who work in our operating rooms every day are the real heroes of this work. Rather than resist, they welcomed REAL in hopes it will help all of us provide safer care of the children entrusted to us by their parents. The Waltham operating suite is a special place. The anesthesiologists and nurse anesthetists; surgeons, surgical fellows and physician assistants; circulating nurses, scrub nurses and techs who "exposed" themselves to open review and analysis are true professionals as defined as having "qualities of skilled and educated people such as effectiveness and seriousness of manner". (https://dictionary.cambridge.org/us/dictionary/english/professional)

My MMSc-Medical Education never happens without others helping me find the time to engage in the coursework, go to the required classes, and do my homework on time. My orthopedic and hand surgery professional colleagues (adapting their research, educational and clinical care schedules); ambulatory administrative staff (starting full clinics at 6am three days a week); and our orthopaedic leadership and management team (meeting with me at times and in buildings they never knew existed) at Boston Children's were flexible, understanding and I think a little bit amused. Amazingly, we got it done. Joanne Casey was spectacular. Without her, I would still be at step one of this process. Incredibly, my clinical and surgical volume was as busy as ever and I met my leadership obligations in the department, residency program and hospital. It was not seamless but it was effective and important. I am grateful and indebted. You are amazing.

Finally, what can I say about a wife (Janet) and now adult children (Rebecca and James) who support me as I journey these roads of inquiry? Simply put, I am off the charts in love with you.

"This work was conducted with support from Students in the Master of Medical Sciences in Medical Education program of Harvard Medical School. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard University and its affiliated academic health care centers."

Chapter 1: Background

Overarching Problem Statement: Our ultimate goal with the REAL project is to increase the safety and quality of surgical procedures for our patients. More specifically REAL was designed to achieve a next step in increasing safety, namely to identify in real time what is preventing us from achieving reduction, or even better, elimination surgical error (*Reason 2000*). Traditional surgical training and education have been predominately designed to (1) increase the acquisition, retention, and application of core knowledge necessary to deliver the highest level of care for each health condition; and, (2) foster the development and distribution of necessary technical expertise that keeps pace with, or even leads, innovation over the course of a career. Similar goals are true for training programs among other professional members of the operating room team including scrub technicians, nurses, physician assistants, residents, fellows and anesthesiologists. However, there is evidence that indicates errors during procedures are as much due to performance gaps among of the non-technical components of leadership, communication, teamwork, and situational awareness as it is to decision making based on knowledge content and technical skill execution. Unfortunately, despite the prevalence and consequences of non-technical errors, less time overall is devoted to fostering and learning the behaviors of teamwork such as leadership and communication in surgical training and continued medical education. This is in stark contrast to the abundance of theory and evidence supporting that in highly complex situations, such as a surgical procedure on patients under anesthesia, high performing teams create less risk for themselves, for their institution, and most importantly, for the patient.

Case: Wrong Site Surgery. Unfortunately, recent analysis estimates that 25% of orthopedic surgeons will still perform wrong site surgery in their career (*Santiesteban 2016)*.

21% of hand surgeons have already (*Meinberg 2003*). 50% of spine surgeons have performed wrong level surgery (*Santiesteban* 2016) even after implementation of SMaX (Sign, Mark, Xray) processes *(NASS, North America Spine 2016).* Despite mandates, universal protocols, and regulatory reviews, wrong site surgery errors continue at an unacceptable rate (*Helmrecih 1996).* The "burning platform" has yet to be transformational (*O'Connor 2014).* With an annual estimated 330 million surgical procedures performed globally, and 26,000 procedures per 100,000 lives in the United States (US), (*Lancet)* there is clear continued risk to patient, healthcare professionals and organizations. Why?

Strategies to reduce risk have included:

Checklists: Wrong Site-Wrong Procedure-Wrong Patient (WSPEs) became an urgent "burning platform" as of 2008 (*Kotter 2008*). The logic was clear: decrease variation by standardizing the steps for each patient by every surgeon and operating room team from the start to end of each surgical procedure and preventable error will be avoided. (*Gawande 2003, AHRQ 2019 safe surgery*).

The Canadian Orthopedic Association (COA) introduced a Sign Your Site program in 1994; the American Academy of Orthopedic Surgeons (*AAOS 2003*) adopted the same in 1998; and the Joint Commission on Accreditation of Health Care Organizations (JACHO) mandated utilization of the World Health Organization (WHO) universal protocol in 2004 (*WHO)* while defining wrong site surgery as a "never event." In 2004, the American Operating Room Nurses (AORN) association developed a correct surgery toolkit (*JAORN*). WSS and Wrong Site-Wrong Procedure-Wrong Patient (WSPEs) became a focal point in all healthcare organizations where procedures were performed (*Kotter 2008*). The logic was clear: decrease variation by standardizing the steps for each patient by every surgeon and operating room team from the start

to end of each surgical procedure and preventable error will be avoided (*Gawande 2003, AHRQ 2019 safe surgery*). Adapting error prevention processes from well-established aviation safety industry checklist standards for takeoff, flight, landing and taxing, the WHO Surgical Safety Checklist (Figure 2) defined simple steps that are measurable and usable in any surgical procedural setting: *Sign-in, Time-out, and Sign-out (Clay-Williams 2015).* Each process event occurs at prescribed times during the procedure with required sub-steps. Sign-in is performed when the patient, surgical, anesthesia and nursing teams enter the room before induction of anesthesia. *Sign-in* requires confirmation of patient identification, surgical site marking (*Pikel 2014)*, surgical consent, patient allergies, and anesthetic plan. The patient and/or their parents need to affirm the plan before any mild altering medications are given, not only verbally but in writing *(DiGiovanni 2003). Time-out* has to occur before incision. All members of the team must again confirm surgical site marking, matching consent as well as affirm operative plan, expected duration, possible difficulties expected, equipment needed, imaging verification of operative site and needs during procedure, blood products required, and antibiotics given. *Sign-out* occurs at the end of procedure when the surgical drapes have been removed and includes final description of procedure performed along with review equipment or patient safety concerns that could be improved. Adaption of the surgical checklist has been cited to reduce preventable errors by up to 50% as evidenced by the Safe Surgery Saves Lives study group comparative pre-checklist and post-check-list outcomes in eight hospitals in eight countries globally for non-cardiac surgery adult patients (*Haynes 2009*). Analysis of other large cohorts pre- and post-checklist implementation indicate similar short-term reduction in preventable errors as cited by several authors. (*de Vries* 2010, *Bergs 2014, Geraghty 2017, Lewis 2012, Mascherek 2013*, *Newman-Toker 2009*).

Addressing Human Error: These processes adhere to the understanding that people are prone to failure, especially when performing work on the "sharp end" (*Cook, Woods 1994*), and processes addressing these factors (of which checklists are one) specifically put in place and adhered to can dramatically lessen the risk of preventable error. For example, Rasmussen's skill, knowledge and rule model of human error (*Rasmussen 1983)* and Reason's "swiss cheese" model of error risk and management were adapted into the universal protocol *(Sarker 2005 , Reason 1990*). Redundancy is built into the system as redundant checks have been shown to exponentially reduce error risk if they are independent checkpoints. (*Kwan 2006*). Unfortunately, despite strong grounding in behavioral science, two thirds or less of potential surgical errors have been deemed preventable by these processes, even with complete adherence to protocol (*Borchard 2012, Greenberg 2007).* That leaves a gap (*Bonnaig 2014.)* Non-compliance with a checklist leaves a broader gap. There are many reasons for non- or incomplete compliance *(Deutsch 2017, Gersick 1990)* but one important issue is the "it can't happen to me syndrome." People can be lulled into believing they are not at risk for a rare but preventable professional disaster. Optimism bias can lead surgical teams to disregard the overall reality of situational risk and believe they are excluded from potential negative outcomes. (*Osbaldiston).*

Incentives, Penalties: One thought has been to use reward or penalty systems to lessen performance risk. If error prevention logic and processes could not completely prevent major error outcomes in surgery, maybe financial penalties could prove to be the necessary change agent. In 2007, the Centers for Medicare and Medicaid Services (CMS) stated hospitals would not be reimbursed for any additional costs associated with WSPEs. In 2009, CMS announced that hospitals will not be reimbursed for any costs associated with WSPEs. (*CMS.gov)* In addition, since 2015 the Hospital Quality Assurance program Medicare pay-for-performance

program penalizes the bottom quartile of performance with reduced payments. (*CMS.gov)* Money matters, right? But to whom? Clearly to the hospital or over-arching academic medical center. But does it matter all the way down to the grass roots level professionals working in the operating room? Does an error impact the compensation for members of the operating room team? In most healthcare organizations, the answer is no.

Similarly, reward systems could be a difference maker, but hospitals and professionals get paid for a wide range of outcomes. Most healthcare organizations do not have additional specific rewards for positive outcomes, let alone for compliance of a checklist. High performance is expected, and often assumed, but compensation is rarely impacted. Of note, detailed analysis of performance is often lacking so individuals and teams are unaware of how they are performing compared to peers within and across organizations.

Transparent Reporting: What about transparency in reporting? Professionals, especially physicians, surgeons, and nurses are motivated by high "grades"; years of experience in rank order schooling has fostered drive for excellence in professional evaluations. Internal motivation to prevent failure is real and is one trait that defines high level professional behavior. Exposure of less optimal results compared to peers can also drive change. Fear of liability can influence some behaviors but not necessarily to a higher quality, safer outcome. Failure to report required serious safety events can have reputational, financial and accreditation consequences for an institution. (*Lingard 2006)*

Successful Outcomes: The Hospital Quality Alliance reporting theory maintains that through accurate, regular reports, individuals and teams will improve their performance. There is evidence that open reporting can have an impact in reduction of errors as seen through the

hospital readmissions reduction program and the hospital acquired condition (HAC) reduction program (*CMS.gov).* Being watched may help change behavior, at least in the short term. (Hawthorne effect) (*Sedgwick 2015).* As noted previous, evidence indicates that improvement on complete adherence to checklists will lessen the risk of major error, such as wrong site surgery. Further, near misses, minor failures, and even inefficiencies such as due to equipment problems, should diminish too. This concept of preventing minor errors lessening the risk of major errors follows the "broken window" theory of crime prevention. By policing thoroughly all small crimes, larger crimes were prevented *(Wilson1982).* Paying intimate attention to detail, and correcting even minor problems in real time, builds teams of high accountability and performance. In addition, reporting to a central database with shared learning can improve safety of care across organizations.

Compliance Monitoring: Unfortunately, self-reporting is complex and variable in accuracy. The present standard for assessing compliance is completed via professional audits. After an initial reduction in safety risk with the implementation of WHO checklists, safety risk has nonetheless remained at an unacceptable level for patients, professionals, and institutions with published compliance rates between 62-92% (*Johnston 2009, Russ 2015, Sparks 2013, Stahel 2014).*

Reasons cited for less than 100% compliance and increased risk of error due to non-compliance include: (1) lack of knowledge about the importance of the checklist playing a critical role in preventing error; (2) incomplete or inaccurate checklist performance impacting professional evaluations or compensation; and (3) limited harmonization of checklist acceptance and compliance among OR team members (*Fourcade 2012, Sewell 2011, Singer 2016)).* Missing

data points are common, making analysis to guide effective improvement difficult. There are times when professionals miss the major difference between "work-as done" and "work-as-imagined" or "work-as-documented" (*Deutsch 2017, Sendlhofer 2016.)* Getting this work done right, not just imagined or reported as such matters. Over-reporting of adherence to guidelines is known with self-reporting rates exceeding objective rates of guideline adherence by a median of 27% in an analysis of 326 studies of guideline adherence. (*Adams 1999*) Thus, reporting bias is a concern.

High Reliability and Safety: High reliable organizations embrace a culture of safety that adheres to (1) open recognition of high-risk work; (2) commitment to be consistently safe in operations; (3) a blame free environment that leads to fearless self-reporting of errors and near misses; (4) collaboration across disciplines and amongst individuals to enhance safety and resolve impediments to safe care; (5) data driven quality improvement; and (6) organizational commitment with resources to address safety concerns (*AHQ PSNet 2019 culture of safety*). High reliability organizations focus on failure risks, are resilient when (not if) errors occur, hold themselves accountable, and work tirelessly and systematically to learn and improve (*Edmondson 2003, Rosenbaum2019).* High reliability understands not only that *to err is human,* but that *to err is inevitable.*

Resiliency: Hollnagel et al (*Hollnagel 2015, Hollnagel 2017)* evaluated error prevention, safety management strategies and introduced the concept of a spectrum from *safety 1* (the absence of accidents or "freedom from unacceptable risk") to *safety 2* (the ability to succeed under varying conditions). Safety 1 was described as looking for what can go wrong and success defined as no adverse events; Safety 2 is looking for what is going right and managing beneficial

performance variability. One of our observational goals in REAL to see what happens every day, not only in terms of increased risk, but also productive behaviors that can guide meaningful change *(Flin 2007).*

Roles, Responsibility and Micro-climates: A major issue with team performance is ultimate responsibility for specific tasks, shared tasks, and in the end, outcome. The ABC's of "assess, blame, criticize" can emerge in the aftermath of a crisis. In a hierarchy management structure, the "buck stops here" or "I'm in charge" has some advantages in delineating responsibility. There are plenty of times this perspective still predominates operating room behavior (*Lingard 2012, Lingard 2005).* Unfortunately, do we clearly know at any given moment which person is charge of what elements of the event? Silo behavior and isolated environmental microclimates in the OR may put the patient at risk (*Roussin, 2008, Roussin 2014).* It all goes fine until it doesn't. Then what? There is ample evidence that high performing teams who practice, do better than even high-level professionals without proper support in complex endeavors (*de Rond 2010).* Surgery is clearly a complex endeavor, requiring all members of the team to be fully engaged and collaborative (*Cooper 2018, Raemer 2016).*

REAL seeks to understand every day moments in the operating room that may translate into greater or lesser risk to patients. These include the above-mentioned elements such as incomplete use of surgical safety checklists and inattention or poor collaborative communication during key diagnostic moments; for examples, anesthesia induction, fracture pinning, fluoroscopy review, or any critical part of a surgery that may lead to premature diagnostic closure or confirmation bias. Additionally, REAL also seeks to find the safety 2 moments and behaviors that guide and determine positive outcomes. The Standard Operating Procedure (SOP)

that emerges within this study will provide the start of a "REAL 'service-line' blueprint" for future adoption in other procedural spaces in our institution and/or other hospitals built on the foundation of current knowledge and additional understanding developed through this project. The blueprint will be process-driven and adaptable, potentially as an initiative to leverage live audio-visual recorded events; a process that traditionally has been difficult to develop and maintain. In addition to the REAL quantitative methods analyzing compliance, qualitative analysis with confidential post-event debriefings and interviews were performed to bring new insights into the professional barriers to complete compliance; as well as positive adaptive behaviors that enhance better team communication and leadership skills.

In summary, our aims for REAL are to (1) design, iterate and develop standard operating procedures and successful adoption of audio/visual recording and analysis of live events in our pediatric operating rooms in order to (2) discover, study and solve barriers to complete compliance that will enable us to approach "never event" safety in the procedural areas with WSPEs and other preventable errors as well as (3) discover, study and reinforce positive, maybe unrecognized ongoing professional behaviors and interactions that create safer care environments.

Chapter 2: Data and Methods

2:1 Introduction

REAL required many iterative steps from study design and implementation to data acquisition and analysis. The concept of live analysis of surgical team performance by in person and separate audiovisual analysis *(Jacobs 2010, Jiang 2010, Parker 2014, Rowlands 2014, Taylor 2011, Wurster 2017)* may seem easy to some; but, getting a healthcare institution and individual professionals to "expose" themselves to complete, unedited review of performance required deep commitment to betterment and trust of the process and individuals involved. In my opinion, just getting REAL done at all is remarkable.

Sequentially,

(1) DESIGN: pre-implementation analysis and study design with consultation of multiple content matter experts (CME);

(2) TOOLS: choice of NOTSS, ANTS, SPLINTS validated instruments (*Fletcher 2003, Flin 2010, Flin 2006, Mitchell 2013, Yule 2018, Yule 2008, Yule 2016, Yule 2013)*. Development of additional compliance checklist, OR team task performance scoring and debriefing tools;

(3) LEGAL: institutional legal, executive level, institutional review board and quality officer input and approval of project as quality improvement with defined patient/family consent process and management of audiovisual recording tapes;

(4) FUNDING: achievement of grant funding from CRICO;

(5) SOCIALIZATION: introduction, socialization and participation acceptance by nursing, anesthesia and surgical staff. Key step in engendering psychological safety;

(6) SCORER TRAINING: training of scorers (five) on compliance checklist, OR team task performance, NOTSS, ANTS, SPLINTS for in OR and out of OR audiovisual scoring;

(7) PILOT: Phase I pilot assessment of feasibility with 3 OR days, 11 surgical cases of varying complexity by single surgeon, same nursing team and variable anesthesia teams. Coding by 3 “in-OR” and 2 “out of -OR” scorers with comparative analysis of Phase I surgeries;

(8) DEBRIEFING: formalization of debriefing process with modified delphi method development of debriefing scoring tool;

(9) QI 1: use of rapid cycle improvement (RCI) interventions on audiovisual technology acquisition and analysis along with further development of scoring tools;

(10) EXPANSION: Phase II REAL study with four surgeons (2 OR days each, total 24 operations), same facility site as pilot phase, variable nursing and anesthesia teams that adhered to standard operating procedures for case and work assignments without regard for REAL. Scoring by 3 rotating in OR compliance coders, the same 4 out of OR audiovisual team performance coders and same single debriefing coder with qualitative analysis expertise;

(11) ANALYSIS 1: analysis of comparative data after Phase IIA of 4 OR days, one day each by each surgeon, total 12 operations, variable anesthesia and nursing teams;

(12) QI 2: modification of audiovisual scoring protocol through second modified delphi method;

(13) ANALYSIS 2: analysis of data of Phase IIB surgeries (same 4 surgeons, one operative day each, total 12 surgeries, variable nursing and anesthesia teams).

2.2: Materials and Methods Introduction

REAL is a mixed methods, quality improvement study.

(*Nicolini 2011)* intended to: a) further discover root causes of team-behavioral challenges and non-compliance with safety standards in the pediatric operating room environment and b) design solutions to the identified challenges.

REAL was implemented successfully over the course of one year in the Boston Children's Hospital Waltham Surgery Center. Prior to the start, the hospital legal department reviewed the project design in detail, met with project leaders, made recommendations for changes that resulted in (1) an agreed upon consent for patients and families and (2) a policy of participation for hospital professional staff. The hospital Institutional Review Board leadership reviewed all documents, met with us several times in person and electronically, made suggested changes, and both recommended and approved REAL as a quality improvement project and study. The hospital Vice President of Quality and Safety similarly approved and supported REAL as a quality improvement project. An agreed upon policy of review, storage, and eventual destruction of the audiovisual tapes was an important element of eliminating concerns about employees and patients recorded being a vulnerable population requiring appropriate IRB and hospital protection. The goal was for REAL to be similar, but clearly more complex, than hand washing requirements for all employees that make care of our patients safer. The executive leadership of the hospital (CEO, COO, Surgeon in Chief, Physician in Chief, Chief Nursing Officer, Chief Anesthesiologist) endorsed REAL. CRICO, the Harvard indemnity insurer, awarded the REAL principle investigators two years grant funding from a competitive solicitation.

REAL is comprised of three major components: (1) real time audio-visual recordings and in OR assessment of live surgical events; (2) structured scoring and analysis of: a) adherence to

safety-oriented compliance protocols and processes and b) non-technical (teamwork) behaviors exhibited by the team and team members; and, (3) structured debriefing conversations

Through formal presentations at faculty meetings, departmental and division grand rounds, operating room staff meetings as well as regular discussions with Waltham operating room staff and leadership, REAL was socialized, concerns solicited and addressed, before implementation

2.3 Data Acquisition

2.3a Audiovisual Recording, Storage and Analysis

The audiovisual recording set up was designed to be (1) of high quality; (2) a simultaneous recording of all activity in the operating room from "wheels in" to "wheels out"; and (3) unobtrusive to lessen the Hawthorne effect. Adaptive changes in audio recording, such as gain control and microphone placement, were made by a simulation engineer throughout the day as needed. At the completion of the surgical day, the de-identified files for each surgical case were loaded and temporarily stored on the CAE Intuity LearningSpace for purposes of coding review.

Each surgical procedure was recorded with three Axis M3046-V Network Cameras mounted with Gorillapods around the room, and a feed from the STERIS system, typically the light camera or scope view for arthroscopic live recording. The positioning of cameras is determined prior to the surgical day, and, with the exception of the STERIS views which was modified by the surgical team, and mounted unobtrusively. The distribution of cameras captured all activity in the room, specifically nursing (circulator and scrub including change of staff),

anesthesia (nurse anesthetist, anesthesia faculty, including change of duty communication, direct visualization of anesthesia machine and monitoring equipment), and surgical (attending, fellow, resident, and/or physician assistant) teams during all times the patient was in the room.

Audio recording was done through four microphones: two Audio-Technica Pro45 ProPoint® Cardioid Condeser Hanging Microphones and two Audio-Technica ATW-T1006 Boundary Microphones. One hanging microphone was positioned towards the anesthesia team and the other towards the surgical team. Positioning of the boundary microphone varied, but was most often placed sterilely within 2 feet of the surgical team and scrub table. In certain cases, one boundary microphone was placed on the anesthesia machine. Placement of boundary microphones was dependent on the surgical procedure and quality of audio from an area of observation.

Optimal camera placements throughout the OR were assessed and implemented in the morning before the operating team entered and began set up. The camera placements were not altered once recording commenced. Video recordings ran directly into and were reviewed via CAE Healthcare Intuity LearningSpace on an Apple MacBook Air. Audio recordings passed through an RDL RU-MX4 Professional 4 Channel Microphone/Line Mixer. Audio quality was evaluated in real time and a simulation engineer made any possible changes as was necessary. Audio was then fed into CAE Healthcare Intuity LearningSpace and synced automatically with the visual input. All de-identified cases were temporarily stored on CAE Healthcare Intuity LearningSpace in order for coding to occur. Each surgical case was saved as an individual file that could be accessed only by assigned Boston Children's Hospital coders via secure password log-in.

Iterative improvements in the quality of the AV systems and capture were achieved from Phase I into Phase II. Regular electronic and in person meetings were held to maintain quality assurance of the AV capture and analysis.

2.3b In OR Scoring Day of Surgery

Each surgical day had one of three rotating trained in OR live scorers present for the entire day, participating in a total of 3-4 days each. All three are non-conflicted professionals not participating in the care of the patients and not members of the professional care teams involved. They were specifically trained to assess compliance and, most importantly, team communication and leadership behaviors with validated and reliable quantitative coding methods. These professionals are presently engaged in simulation programs at the hospital that serve to improve care and safety by enhancing the technical and non-technical skills of our procedural teams.

Completion of the *Compliance and Task Management Checklist* (Appendix 4) for each surgical case was the primary responsibility of the in OR coder and observer. Demographics were obtained regarding surgical site, date, surgeon, case order for the day, case complexity and evaluator. The compliance checklist tool was separated into 11 items and scored by each subcategory within the 11 items: (1) *pre-operative huddles* in terms of participation by various diads and the triad of nursing, anesthesia and surgeon *(Markary 2011);* (2) *pre-operative huddles* in terms of surgical plan, needed equipment, patient condition, associated medical conditions, anesthesia type and timing; (3) *sign-in* as defined by WHO checklist and BCH policy of patient ID, procedure, surgical site marked, weight, known allergies and anesthesia plan; (4) *surgical start timing* in terms of on time or delayed; (5)(6)(7)(8) *time out* as defined by WHO checklist and BCH policy of team introductions, surgical procedure details, surgical site ID marking,

consent matching, operative possibilities, expected duration, imaging if needed, equipment required, fire safety, blood if need, antibiotics given, allergies, weight, tourniquet pressure if used, safety concerns, and a specific call out to speak out if any issues arise; (9) *intra-operative essential steps* for pace of set up, tourniquet times closed loop communication, fluoroscopy imaging review assessment if performed, and number of times circulator left the room for equipment; (10) wound closure for inspection and correct count status; (11) *sign-out* as defined by WHO checklist and BCH policy of name of procedure, any concerns regarding patient or equipment, anything the team could have done better and, (12) open comments for thematic review.

In addition, as a cross check on reliability of NOTSS, ANTS, and SPLINTS (Appendices 1, 2, 3) audiovisual recording, in Phase I and II each in OR evaluator also scored each surgical, anesthesia and nursing teams by these validated tools of non-technical performance of task management, leadership, communication, teamwork, and situational awareness.

2.3c Same Day OR Team Debriefing

The in OR evaluator for both Phase I and Phase II surgical days led a 20- to 30-minute adapted advocacy-inquiry-based (*Rudolph 2007)* debriefing with all the OR nurses, scrub technicians, anesthesiologists, nurse anesthetists, surgeons, surgical fellows, and surgical physicians assistants that participated in the surgical cases for that day. With eye on "ergonomics" of embedding debriefing methods among busy OR teams, debriefing times and approaches were tailored from typical simulation team exercises. The debriefing was semi-structured to enhance standardized across evaluators and operating room teams, but open ended to capture themes and concerns unique to the day.

In Phase II, each debriefing was observed and scored by a debriefing tool (Appendix 5A and 5B) with ample space for documenting open-ended comments for later analysis by a qualitative analysis expert. The goals for debriefing were set by group through modified Delphi method to include team enculturation, problem identification, problem solving, and to provide closure to the day in a positive, productive way. The quality characteristics of the debriefing evaluated were: team engagement as evidenced by balance, depth, relevancy, tone, pace and solution-orientation of the discussion.

2.3 d Audiovisual Scoring

All AV coders were trained in NOTSS, ANTS, and SPLINTS validated tools. (Appendices 1, 2, 3) The first five coders were taught and approved by the Brigham and Women's STRATUS Center for Medical Simulation team in an independent course similar to training they do for the American College of Surgeons (ACS). These five coders were involved in both Phase I and Phase II scoring. Two additional coders added for phase II and underwent in house training with BWH STRATUS and ACS videos and inter-and intra-observer reliability testing.

In Phase I, 3 coders reviewed tapes and scored the teams by NOTSS, ANTS, and SPLINTS on the prescribed 4 point scale based on the elements outlined for each category. Grades of 1(poor), 2 (fair), 3(good), and (4) excellent, for each category within each tool were recorded by all coders independently. (*ANTS*=team working, situational awareness, task management, decision making; *SPLINTS*=situational awareness, communication and teamwork, task management; *NOTSS*=situational awareness, decision making, teamwork and communication, and leadership).

In addition to the standard 11 categories for scoring, 4 additional categories and elements were added for a total of 15 elements by separate scoring analysis: task management to NOTSS, leadership to ANTS and leadership and decision making to SPLINTS. These 15 categories were scored separately in addition to the standard scoring tools. The rationale was that in the 21st century high volume and, high complexity operating rooms, required additional categories and elements for each nursing and anesthesia team to be a highly functional member of the triad. It did not make sense to us that neither anesthesia or nursing would have a "leadership" role evaluated. This by structure of evaluation defines the surgeon as the sole leader of the team. Similarly, we felt decision making was required and should be evaluated for nursing and anesthesia too. Further, it was our concern that the role of the circulating nurse was underappreciated and not being evaluated in the SPLINTS and the diads of circulator to anesthesia, circulator to surgeon, and circulator to scrub were vital and should be evaluated through decision making and leadership additional categories. Finally, surgeons clearly have to manage tasks during an operation so we added that category to our amended NOTSS.

In Phase II, 4 coders scored each video independently. (Figure 2) Phase IIA involved scoring from the NOTSS, ANTS, SPLINTS handbook directly. There was some mild variation between the two original coders and the two added coders. A discussion was held with a non-conflicted analyst who fostered additional training and development of a more precise checklist that was used in Phase IIB.

Each team was scored by overall team performance (Appendix 6) on a scale of 1-5. (1= notably lacking in most aspects; patient safety seriously compromised; 2= notably lacking in some aspects; patient safety compromised somewhat; 3= neutral performance; acceptable patient

safety; 4= notably excellent in some aspects; enhanced patient safety; 5= notable excellent in most aspects; a model of patient safety for others).

In addition, to further compare in- OR versus out-of- OR AV scoring, each AV coder also completed the *Compliance and Task Management Checklist* for each case.

Finally, thematic analysis was used to summarize, evaluate and categorize themes written by all 5 scorers (in-OR (1) and out-of- OR AV (4)) on their compliance checklist and subspecialty tool kit scoring sheets.

2.3e Statistical Methods

Data was imported to a statistical analysis package and "cleaned" in order to identify incomplete, inaccurate or missing data. Corrections were made where possible. The two data sets, NOTSS/ANTS/SPLINTS (NAS) and OR compliance (COMP)were merged to allow for comparison of the different scores.

A variety of descriptive statistics as well as parametric (e.g. t-test and pearson r correlation) and non-parametric (e.g. Wilcoxon rank-sum test, Fisher's exact) statistical tests were used to describe characteristics and assess differences between various groups and subgroups. SAS 9.4_ was used for all quantitative analysis.

Thematic analysis was used to identify recurrent themes and patterns from written (open-ended) responses resulting from both project rating sheets and debriefing assessment sheets. This analysis was completed using manual methods (primarily debriefing sheets) for themes

from project rating sheet notes. This analysis was completed using paper cluster analysis and theme patterning.

2.4 Results

Feasibility, Institutional Endorsement: The Phase I pilot started with volunteer nursing scrub and circulator as well as anesthesia faculty and nurse anesthetist participation. By the end of Phase I and all through Phase II, case and room assignments for anesthesia and nursing staffing REAL surgical days followed usual working protocols. The surgical three days of Phase I were performed by the PI. The 8 surgical days of Phase II were performed by the PI and three additional orthopaedic surgeons on randomly selected typical work days for all four surgeons. No anesthesia, nursing, or surgical staff refused to participate.

Patients and families were consented for each procedure. No patients or family refused participation (35 operations).

REAL was accepted as a quality improvement project, implemented and executed without disruption.

Feasiblity: Audiovisual Recording, Storage, Review and Analysis

The three coders in Phase I scored all cases using NOTSS, ANTS, SPLINTS, and the compliance checklist. Advised changes in audiovisual equipment quality and placement were made to improve the reliability of recording and scoring. Since Phase I was designed to be a pilot, this was expected.

The four coders in Phase II were able to score all elements of NOTSS, ANTS, SPLINTS as well as the compliance checklist for each case. There were a few cases in which obstructions to audio and visual quality due to typical equipment use or maneuvering in the room occurred. For example, a "just in time" adaptive change was necessary during the initial shoulder arthroscopy case due to proximity of the suction to the microphone. Following this initial case, modifications were made to improve the audio set up. There were a few cases where the audiovisual recording started after the patient entered the room; however, patient transfer to the surgical bed was recorded in every case. Sometimes the recording files would "freeze" during review by the coders and the visual input would decouple from the audio. The solution of restarting the LearningSpace system frequently resolved this problem. Occasionally, the AV technicians had to reconfigure the connectivity of the audio to visual on the Intuity LearningSpace, which resulted in a delay of scoring of that particular case by a day or two. The four coders communicated any difficulties electronically to the technical team and copied the other coders to increase efficiency of resolution and scoring. Despite these challenges, we experienced success with this platform and process during REAL.

The review and scoring of 3-4 surgeries on each day was time consuming. Total in surgery time was 4-8 hours, with turnover and set up time adding to a total of 10-12 hours at the site. Four coders, along with one in-OR evaluator, were used to test both reliability and accuracy of the assessment tools. In the future, 5 scorers of complete recordings of every operation will not be needed based on statistical results that indicated reasonable inter- and intra-observer reliability. Also, outcomes from analysis of video scoring data, ie pearson r correlation and regression modeling, appear to more accurately reflect the associations between NOTSS, ANTS, SPLINTS scores and compliance scores. Thus, live in OR scoring is not needed.

Surgical Demographics Phase I consisted of 3 surgical days and 11 surgical cases by a single surgeon (>25 hours of surgery). Phase II consisted of 8 surgical days, 24 surgical cases by 4 surgeons, consisting of >50 hours of direct and audiovisual observation each. All of the surgeons in REAL were orthopaedic surgeons, specializing in hand, sports, and traumatic injuries of the upper and lower extremities. The surgical cases were single limb reconstructions of low to moderate complexity: (4) complexity class 1, (17) complexity class 2, and (3) complexity class 3 operation. There were 13 arthroscopy reconstructions (6 knee, 1 hip, 4 shoulder and 2 elbow); 8 elective hand or foot reconstructions; and 3 elbow or forearm reconstructions. All patients were American Society of Anesthesia (ASA) physical status (*American Society of Anesthesia)* 1 (healthy) or 2 (mild systemic disease). By hospital policy, no class III or higher patients can have surgery in the Waltham facility. Therefore, this was a preselected healthy to moderately healthy patient population with single limb musculoskeletal disorders requiring surgical intervention.

In Operating Room Compliance Checklist Scoring (Table 2)

The in Operating room pre-surgery huddles were often observed between surgeon-nurse (15/24, 62.5%), but surgeon-anesthesia (6/24, 25%) and nurse anesthesia (8/24, 33%) infrequently after the patient entered the room. Interestingly, often these huddle conversations were noted to occur during the concluding part of the prior surgery to review surgical plan, needed equipment, and patient condition. In addition, these observations may be lower than actual as coders and OR team did not have access to huddles that may have occurred in the pre-operative area, hallway or staging rooms outside of the OR.

Complete patient sign-in's upon entry into the operating rooms were performed infrequently. (8/24, 33%)(table 2) Incomplete sign-ins were common. Of the 6 sign-in items, in the incomplete sign-in cases, there was no sign-in at all in 3 cases; 2 cases each had 1 of 3 items; 3 cases had 4 of 6 items and 6 cases had 5 of 6 items. There were no differences across surgical cases in terms of defined complexity, type of surgery, and surgeon, surgical team or day of surgery in terms of compliance scoring.

Time outs were often performed but frequently incompletely so. All observed surgeries had six or more of the required items but none had all items of the time-out completed. (Table 2). The elements performed most consistently were surgical site identification and marking (100%). The components most often missed were encouraging team members to speak about safety concerns during surgery (10/24, 42%), fire safety (10/24, 42%), and imaging review (10/24, 42%).

Sign outs were rarely complete (only 1 of 24 observed all three elements) and frequently not performed at all. (Table 2)

In terms of essential steps in the OR, communication around *tourniquet times* demonstrated that appropriate readback-feedback occurred 100% of the time at 60 and 90 (10 and 4 minutes respectively) There were two cases where the tourniquet time was longer than 120 minutes but not 140 minutes and one of the two cases did not have readback-feedback. *Fluoroscopic review* was done in every case when fluoroscopy was a part of the surgical procedure but most often was a single provider review (71%) and not a collaborative team review with input (29%). *Equipment issues* were limited with 4 % of cases requiring retrieval.

Wound closure details were performed 80% of the time. Complete count communication was 50% of the time.

Remote Audiovisual Scoring All categories based on prescribed elements for NOTSS (situational awareness, decision making, leadership, teamwork and communication), ANTS (team working, situational awareness, task management, decision making), and SPLINTS (situational awareness, communication and teamwork, task management) were scored for each and every case by all four coders (Table 1). There were no incomplete scores. The 4 categories for NOTSS, the 4 categories for ANTS, and the 3 categories for SPLINTS were combined for each system to create an aggregate score for each professional subgroup of the operating room team. The overall combined aggregate scores in REAL scores were for ANTS (290), NOTSS (341), and SPLINTS (232). When these scores were adjusted for number of variables, there were statistical difference between professional subgroups ($p<0.001$). (Tables 4 and 6) The anesthesia team scored lower statistically than the surgical and nursing teams in the validated NOTSS, ANTS, SPLINTS scoring. In addition, when assessing by quartiles, the anesthesia had statistically fewer top quartile scores than the other two teams. (Tables 4 and 6)

Modified (non-validated) NOTSS, ANTS, SPLINTS Scores

The analysis was repeated with the addition of the categories of task management to NOTSS; decision making and leadership to SPLINTS; and leadership to ANTS. The now 5 category system of assessment for each professional subgroup was compared with the validated NOTSS, SPLINTS, and ANTS scores. There were no statistical significant differences (Table 3).

Comparing each professional group with the modified scoring system, there were statistically significant differences in the categories of teamwork and communication, decision making and leadership but not situational awareness, (Tables 4 and 6) with surgical teams with higher scores than the anesthesia and nursing teams.

When aggregated NOTSS, ANTS, and SPLINTS (NAS) scores compared across operative days, (Table 3) we found that there were some minor mixed differences by specific operating room teams for any given day in Phase II but overall, scores for each team were consistent across teams, days, and cases. For total NOTSS/ANTS/SPLINTS (NAS) aggregated scores, there were no significant differences ($p=0.11$). There was a minor difference in situational awareness ($p=0.037$) that could have been by chance, and not for task management (0.12), teamwork and communication (0.37), decision making (0.11) and leadership (0.09).

Comparing case complexity to performance on NOTSS, ANTS, and SPLINTS scores revealed no differences across categories 1, 2 and 3 across surgical days but again, most of these cases were in healthy patients with single musculoskeletal system disorders.

Audiovisual compared to Live in OR scoring (Table 5) The comparison of compliance checklist, validated NOTSS, ANTS, and SPLINTS scoring by live in OR evaluators and by audiovisual recorded showed statistically significant differences. These differences were noticeable as well when associations between NAS and compliance scores were compared: aggregated video scores reflect a positive association between NAS and compliances, in OR scores present a negative association. (Both pearson r positive correlations and regression modeling were used; regression modeling outcomes are reported as the p values are more meaningful for our research). Given video scoring was more focused without distractions,

allowed replay to be certain all the behaviors were observed and recorded accurately, we conclude video scoring more accurate and reliable of NAS and compliance scoring

Performance of Scores on Compliance Checklist to NOTSS, ANTS, SPLINTS Overall aggregated compliance checklist and NAS (NOTSS, SPLINTS , ANTS) scores were a positive association between NAS and compliance score correlations (coefficient =1.2, p=0.044)

Thematic Assessment of Written Comments on NOTSS, ANTS, SPLINTS and Compliance Checklist Scoring sheets.

Through thematic scoring, several interesting and key themes emerged. First, we learned that the OR teams were frequently concerned about **communication**. Communication was consistent as both a challenge as well as a solution across all five debriefing sessions. The more complicated the situation got (two surgical teams, two sequential rooms of surgeries by same surgeon as examples), the more important clear, consistent, detailed communication was as noted by the professional team members' opinions. A particular challenge for communication was "too many people in the room" as noted in 3 of the 5 debriefing sessions.

A **clear leader/manager** was noted to be essential to resolution of any complicated situations; and, to build team unity and care delivery. In addition, clear communication on steps, progress and decisions before and during the procedure was deemed very important and sometimes lacking.

Recording equipment and additional REAL personnel was noted to be mildly intrusive in the beginning of Phase 2, even to the point of an expression of potential decreased patient safety in debriefing #1; but during the last three debriefings, REAL was not even

mentioned as a concern and noted to be normal operating conditions as they "went to work and forgot about it."

Themes that coalesced into change discussion included: (1) the OR is a "tough environment" to manage all the people, patient, family and case complexity; (2) quiet conversations amongst subgroups are frequent and limit valuable information sharing between teams; (3) open, audible communication in the noisy OR environment that engages the entire OR team is preferred and important but performed less frequently than desired; (4) specific detailed communication about each surgical plan to all members of team before the start of the procedure are rare but preferred to lessen lapses of preparedness or understanding; (5) morning huddles would help expediate care, increase preparedness and build team cohesion; (6) the OR does not yet have a full culture of speaking up, as this behavior is too variable at present; and (7) nursing and anesthesia defer to the surgeon's need to focus on task and care delivery.

Summary of Video Recording Themes from NOTSS, ANTS, SPLINTS written comments in scoring boxes and end of form write-ins. The AV scorers mostly noted high level clinical care given by each subspecialty team and by the overall performance of the team, but found gaps that put patients and professionals potentially at risk. No harm was done as they did not observe major or minor errors in these 24 operations. However, they were surprised by the potential risks of incomplete compliance that was more frequent than they expected.

The 4 scorers commented that the Waltham OR teams were gentle and kind to the patients and their parents when present for induction; and, still managed to do their job well under observation while often providing entertainment with distracting song or story. Although

there was no scoring tool for these interactions, they do represent a high level of professional behavior, managing well both the patient's and their parents' physical and emotional needs.

Task management was rarely an issue. Each team did their job well most of the time. However, there were frequent times when anesthesia was not actively engaged in communication with the rest of the team. There was usually clear communication between nurse anesthetists and anesthesia staff when they were taking over for each other during a break, but almost never did the nursing or anesthesia staff announce the change to the surgeon and other scrubbed or circulating team members. *Remaining quiet* while doing their job for anesthesia and nursing seemed to be a common theme.

Sign out was rarely performed, so the evaluators wondered how problems with equipment that occurred in that particular case could be resolved before another case. Although equipment problems were rare, when present there did not appear to be a closed loop mechanism to be certain resolution occurs before the next surgery requiring that equipment.

2.5 Results Discussion

REAL identified a number of behaviors that either add risk or provide safer environments for patients undergoing surgery during the 11 days, 35 cases of observation, evaluation and analysis.

The REAL Study has demonstrated both direct observation and audiovisual assessment of individual and team performance to be feasible with audiovisual analysis more reliable. Audiovisual scoring overall NOTSS, ANTS, and SPLINTS scores to be positively correlated with compliance scoring checklist. The compliance scores were lower than expected and indicate

risk for error. The assessment tools were useful and appear adaptable to coaching of individuals and teams. *(Gawande 2013, Hu 2017*)

In the *"added risk"* category (Safety 1), we observed: (1) frequent microenvironments of professional behavior focused only on their specific task management; (2) limited open professional communication between sub-teams that fosters learning, collaboration, and attentiveness across teams; (3) frequent incomplete compliance with time-outs; (4) frequent partial to non-compliance with both sign-in and sign-out; (5) checklist time out on wall placard in OR appeared to be prone to cognitive overload and lacked defined roles for all members of the professional team; (6) normative culture of anesthesia and nursing appeared to be to remain silent during surgery unless spoken to by surgeon (granted no major anesthesia complications occurred in this study); (7) patterns of behavior that are not compliant with hospital policy (checklist protocols) continued without commentary or correction; (8) assumptions of "no risk" that may have been influenced by healthy status of the patients and the familiarity of the staff in the environment appeared to add risk; (9) limited communication on change of staffing was of concern regarding team familiarity and awareness; and (10) there was no observed closed loop communication and resolution on challenges encountered with equipment.

If any or hopefully all of these are improved upon, there will be less error risk and potentially safer outcomes.

In the "*safer environment* "observation category (Safety 2), we observed: (1) repeat examples of efficient whole-team pre-operative huddles for the next case occurred toward the conclusion of the prior case occurring while the entire team was present; (2) pre-emptive preparatory discussion regarding more complex cases, patients, and families early in the day to

prepare the team; (3) collaborative discussions within surgical team (attending-fellow-resident-physician assistant), anesthesia team (attending-nurse anesthetist), nursing (circulator-scrub technician/nurse) around decisions and during critical moments; (4) rotating collaborating diads throughout a surgical case and day including: (a) circulator-surgeon before patient in room; (b) scrub-surgeon before prepping and draping; (c) circulator-anesthesia during patient coming in the room and transfer to bed; (d) surgeon-anesthesia during positioning; (e) surgeon-scrub during surgery; (f) circulator-fluoroscopy technician during room set up; scrub-radiology technician during sterile draping of C-arm; (g) surgeon-fellow during critical parts of surgery; (h) anesthesiologist-nurse anesthetist during induction, change of staff, and extubation; (5) emphatic care of parent and child during the safe, efficient induction of anesthesia; (6) predominately professional tension free respectful interactions amongst staff; (7) commitment to professional betterment and improved safer care of the patient by nursing, anesthesia, and surgical staff as observed in debriefing discussions; (8) kindness and cooperation amongst staff.

These safety 2 behaviors have a great deal to do with safe, high quality care of the patients that we observed.

Chapter 3: Discussion and Perspectives

We have learned so far that:

(1) Audiovisual assessment of professionals and teams is feasible (and possibly a more accurate and reliable option to in OR direct observation);

(2) Non-technical skills assessment of operating room teams reveals important findings relevant to teamwork and patient safety;

(3) Direct monitoring for compliance of checklists to reduce risk and improve safe outcomes for patients is necessary;

(4) Improved methods of "crowd sourcing" compliance behavior and monitoring is needed;

(5) There are professional behaviors which vary from expected that can positively add high value to patient care and outcome;

(6) Even our best technical surgeons and anesthesiologists who are respected for their professional demeanor and knowledge have room for improvement on compliance, leadership, situational awareness, teamwork, and communication;

(7) Task management was not an issue for anesthesia and surgeons, and rarely so for nursing, in this study cohort;

(8) Our professionals are invested in professional betterment by their debriefing feedback; and

(9) REAL in concert with simulation can create important "cycles of safety" by adapting rapid-cycle improvement industry models of quality improvement directly to in the operating rooms and in so doing - support advances in patient safety (*Rosenbaum 2019).*

3.1 Limitations

REAL is a mixed methods quality improvement study performed in the operating rooms of a hospital satellite facility designed for ASA 1 and 2 patients with less complex surgical procedures than the downtown academic center hospital operating rooms. All the REAL surgeries were orthopaedic sports, leg/foot and arm/hand operations. The lessons learned here will have to be transferred to more complex surgeries in patients with higher co-morbidities across multiple surgical subspecialties.

There is concern of a Hawthorne effect skewing scores to improved performance evaluation. To lessen this risk, we tried our best to have the in- OR evaluators and audiovisual technicians be unobtrusive, quiet and disengaged from the clinical activity. Our goal openly expressed to lessen the risk was for them to be "wall flowers" or even "paint on the wall." We attempted to use small microphones and cameras out of line of site to achieve this. The set up occurred before the anesthesia, nursing and surgical staff arrived for the day.

There is risk of bias having co-PI have roles both as evaluator and study participant. We openly acknowledged the risk, discussed in detail, modified my role from Phase I to Phase II, set up boundary safe guards with open communication for all other participants and established options for opt outs. As with any ethnographic qualitative methodology, recognizing, communicating about, and protecting against bias of the evaluator who is also a member of the observed community is important.

Similarly, my role as Orthopaedic Surgeon in Chief at the Hospital, Professor at Harvard Medical School, and Chair of the Harvard Residency Program has a risk of altering behavior and

communication with our staff, department faculty, other subspecialty staff and faculty, and members of the research team. We openly addressed this often to be certain all participants were comfortable we were protecting each other, the professionals participating, and the data we derived from our study.

There is a limit to our sample size that may affect our and others' interpretation of our data sets.

Finally, by using the validated NOTSS, ANTS, SPLINTS, and then adding the four additional elements, there is a risk of scorer fatigue that could theoretically jeopardize the validated scores.

3.2 Further Research

Expansion of REAL and Evolution of Safety Cycles

The goal in year 2 of REAL and our CRICO grant is to implement interventions identified and developed in Year 1 of REAL including new simulation based training where appropriate, to improve adherence during key safety moments. Our next steps are to (1) inform the Waltham OR teams of our findings in order to make iterative improvement changes based on our analytical data and their input; (2) to build an improved spine surgical team by utilizing simulation based on year 1 results, audio-visual review and analysis of spinal deformity surgery summer 2019, and activate an OR debriefing culture during orthopaedic surgery spinal deformity complex surgeries in the main campus operating rooms; (3) work with surgical executive committee of all chiefs, surgeon in chief, nursing chief, and anesthesia chief to improve OR compliance and safety (subcommittee already formed).

Coaching and Credentialing

The observed and analyzed data from REAL has high value for the individual professional and to the institution in terms of coaching. So often healthcare institutions utilize coaches for physicians with disruptive behavior or who are failing. Aspirational coaching is underutilized. Athletic teams and coaches understand the importance of self-observation and focused practice thereafter. It makes them more ready for "game-time". Similar to athletic coaches, during surgeries surgeon-coaches have to have high situational awareness, make adjustments with their team members, reassess strategies that may not be working, communicate clearly to their team as the manager the task of the operation. In between "games"-surgeries, the whole team with the surgeon-coach should be working on skill development and make iterative adjustments. REAL with NOTSS, ANTS, SPLINTS offer that option and we are utilizing it. We are performing observational scoring and coaching in a Surgical Fellows Leadership Seminar quite effectively including video reviews of surgery by others through simulation performed by fellows and upcoming surgery to be performed live by the surgical fellows with an in OR graders. We are engaged in a separate surgical coaching program across Harvard hospitals through Ariadne labs. We have begun coaching with the 4 surgeons, compliment of nursing and anesthesia teams involved in REAL Phases I and II. This will expand as we move to Phase III in the spinal surgery program this upcoming summer.

In addition, we are looking into how this can impact credentialing. On the two-year cycle of renewal, along with a 360 evaluation, we will pilot using REAL videos.

Ultimately, to leverage data-driven live video capture and structured debriefing as agents of change; and then using these data to inform interventions to improve compliance, hopefully

we can create a self- serving learning cycle of improvement and safety. Furthermore, this cycle will include the impact of live video capture on simulation-based practice, incorporating this tool into simulation training and focusing the simulation debriefing process back to actual surgical procedures. In this feedback loop, live video will inform necessities in simulation, the usefulness of which can then be evaluated by later live video recordings, thus creating an educated movement toward change.

Our new understanding of barriers will enable modifications in processes and training to prevent high-risk errors as the present checklist compliance and team behaviors have performed less than ideal on prevention of error. Additional ultimate goals are (1) a compliance monitoring program that aligns with credentialing; (2) a coaching program that improves important non-technical behaviors; and (3) ideally, to grow a REAL Program/Service into other procedural areas outside of the operating room as a catalyst for reliability and safety enterprise wide. To achieve (3) will require the completion and full vetting of a scalable standard operating procedure (SOP) for sustainable live video capture programs that can be applied to all procedural rooms, such as interventional radiology and endoscopy suites.

Bibliography

Adams A, Soumeral S, Lomas J et al. *Evidence of self-report bias in assessing adherence to guidelines.* I Journal Qual Health Care 1999: (11) 187-192.

AHRQ PSNet 2019 https://psnet.ahrq.gov/primers/primer/5/Culture-of-Safety

AHRQ PsNet 2019 https://psnet.ahrq.gov/primers/primer/18/wrong-site-wrong-procedure-and-wrong-patient-surgery january 2019)

American Academy of Orthopaedic Surgeons Council on Education, Report of the Task Force on Wrong-Site Surgery. Rosemont, Ill American Academy of Orthopaedic Surgeons1998;Available at: http://www3.aaos.org/safety/ewsurgery.htm. Accessed September 1, 2003

Anesthesia Society of America. ASA Physical Classification System. https://www.asahq.org/standards-and-guidelines/asa-physical-status-classification-system

AORN Actions to Prevent Wrong Patient, Wrong Site , Wrong Procedure Events. https://www.aorn.org/-/...site/.../actions-to-prevent-wrong-person-site-procedure-event...

Bergs J, Hellings J, Cleemput I, et al. Systematic review and meta-analysis of the effect of the World Health Organization surgical safety checklist on postoperative complications. *Br J Surg.* 2014;101(3):150-158.

Bonnaig N, Dailey S, Archdeacon M. Proper patient positioning and complication prevention in orthopaedic surgery. J Bone Joint Surg Am. 2014;96(13):1135–1140. doi:

Borchard A, Schwappach DL, Barbir A, Bezzola P. A systematic review of the effectiveness, compliance, and critical factors for implementation of safety checklists in surgery. *Ann Surg.* 2012;256(6):925-933

Clay-Williams R, Colligan L. Back to basics: checklists in aviation and healthcare. *BMJ Qual Saf.* 2015;24(7):428-431.

CMS.gov. Medicare and Medicaid move aggressively to encourage greater patient safety in hospitals and reduce never events. https://www.cms.gov/newsroom/press-releases/medicare-and-medicaid-move-aggressively-encourage-greater-patient-safety-hospitals-and-reduce-never

CMS.gov. Hospital Acquired Condition Reduction Program Fiscal Year 2019 Fact Sheet: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/HAC-Reduction-Program-Fact-Sheet.pdf

Cook RC, Woods DD. Operating at the Sharp End: The Complexity of Human Error. 1994 Research Gate 13: 255-310. https://www.researchgate.net/publication/313407259_Operating_at_the_Sharp_End_The_Complexity_of_Human_Error

Cooper JB. Critical Role of the Surgeon-Anesthesiologist Relationship for Patient Safety. *Anesthesiology.* 2018: 129 (3) ; 402-405. doi: 10.1097/ALN.0000000000002324

Davis WA, Jones S, Crowell-Kuhnberg AM, et al. Operative team communication during simulated emergencies: Too busy to respond? *Surgery.* 2017;161(5):1348-1356.

Dayuta JC, Ong LT, Pang NL, Chia SN, Tan HK, Gomez JM, et al. Process reengineering of preoperative verification, site marking and time-out for patient safety. World Hosp Health Serv. 2013;49(2):4-9. [PubMed]

de Rond, M. Always rely on a team, not individuals. Forbes 2010. https://www.forbes.com/2010/08/05/teams-teamwork-individuals-leadership-managing-collaboration.html#39694b75e86d

DiGiovanni CW, Kang L, Manuel J. Patient compliance in avoiding wrong-site sur- gery. J Bone Joint Surg [Am] 2003;85-A:815-19.

Deutsch E. Bridging the gap between work-as-imagined and work-as-done. Pennsylvannia State Safety Advisory. 2017; 14(2) 80-83.

Edmondson A. Speaking up in the operating room: how team leaders promote learning in interdisciplinary action teams. J Management Stud 2003;40:1419–52.

Fletcher G, Flin R, McGeorge P, et al. Anaesthetists' Non-Technical Skills (ANTS): evaluation of a behavioural marker system. Br J Anaesth 2003;90:580–8.

Flin R, Patey R, Glavin R, Maran N. Anaesthetists' non-technical skills. *Br J Anaesth.* 2010;105(1):38-44.

Flin R, Yule S, Rowley D, et al. The Non-Technical Skills for Surgeons (NOTSS) system handbook v1.2: Structuring observation, rating and feedback of surgeons' behaviours in the operating theatre. Aberdeen (Scotland): University of Aberdeen; 2006.

Flin R, Youngson G, Yule S. How do surgeons make intra- operative decisions? Qual Saf Health Care 2007;16:235–9.

Fourcade A, Blache JL, Grenier C, Bourgain JL, Minvielle E. Barriers to staff adoption of a surgical safety checklist. *BMJ Qual Saf.* 2012;21(3):191-197.

Gawande AA, Zinner MJ, Studdert DM, Brennan TA. Analysis of errors reported by surgeons at three teaching hospitals. *Surgery.* 2003;133(6):614- 62.

Gawande A. Personal best: top athletes and singers have coaches. Should you? The New Yorker 2011 Oct 3. http://www.newyorker.com/reporting/2011/10/ 03/111003fa_fact_gawande (accessed November 25, 2013)

Geraghty A, Ferguson L, McIlhenny C, Bowie P. Incidence of Wrong-Site Surgery List Errors for a 2-Year Period in a Single National Health Service Board. *J Patient Saf.* 2017.

Gersick CJ, Hackman JR. Habitual Routines in Task Forming Groups. *Organizational Behavior and Human Decision Making Processes.* 1990: 47, 65-97.

Greenberg CC, Regenbogen SE, Studdert DM, Lipsitz SR, Rogers SO, Zinner MJ et al. Patterns of communication breakdowns resulting in injury to surgical patients. J Am Coll Surg. 2007;204(4):533-40. [PubMed]

Haynes AB, Weiser TG, Berry WR, et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med.* 2009;360(5):491-499.

Helmreich R, Davies J. Human factors in the operating room: interpersonal determinants of safety, efficiency and morale. Baillieres Clin Anaesthesiol 1996;10:277–95.

Hollnagel E. Prologue: Why do our expectations of how work should be done never correspond exactly to how work is done? In: Braithwaite J, Wears RL, Hollnagel E, editors. Resilient Health Care. Vol. 3. Reconciling work-as-imagined and work- as-done. Boca Raton (FL): CRC Press, Taylor & Francis Group; 2017. p. xvii-xxv.

Hollnagel E, Wears R, Braithwaite J 2015.. From Safety 1 to Safety 2: A White Paper. https://www.england.nhs.uk/signuptosafety/wp-content/uploads/sites/16/2015/10/safety-1-safety-2-whte-papr.pdf

Hu YY, Mazer LM, Yule SJ, et al. Complementing Operating Room Teaching With Video-Based Coaching. *JAMA Surg.* 2017;152(4):318-325.

Jacobs B, Duncan JR, Street M, Murray D. Audio and video recording system for routine documentation of fluoroscopic procedures. *J Vasc Interv Radiol.* 2010;21(5):725-729.

Jiang C, Zhao Y, Chen Z, Chen S, Yang X. Improving cardiopulmonary resuscitation in the emergency department by real-time video recording and regular feedback learning. *Resuscitation.* 2010;81(12):1664-1669.

Johnston G, Ekert L, Pally E. Surgical Site Signing and "Time Out": Issues of Compliance or Complacence. *J Bone Joint Surg* 2009:(91 (11): 2577-2580. doi: 10.2106/JBJS.H.01615

Kotter J. Sense of Urgency. https://hbr.org/2008/08/harvard-business-ideacast-106.html

Kwaan MR, Studdert DM, Zinner MJ, Gawande AA.Incidence, patterns, and prevention of wrong-site surgery. *Arch Surg.* 2006;141:353-358

Lewis BD. Initial evidence: reduced levels of wrong sided surgery [Internet]. c2009.cited 2012 Feb 5. Quebec, CA: Canadian Orthopaedic Association; Available from: http://www.coa-

aco.org/library/practice-management/initial-evidence-reduced-levels-of-wrong-sided-surgery.html.

Lingard L, Reznick R, DeVito I, Espin S. Forming professional identities on the health care team: discursive constructions of the 'other' in the operating room. *Med Educ.* 2002;36(8):728-734.

Lingard L, Regehr G, Espin S, et al. Perceptions of operating room tension across professions: building generalizable evidence and educational resources. *Acad Med.* 2005;80(10 Suppl):S75-79.

Lingard L, Vanstone M, Durrant M, et al. Conflicting messages: examining the dynamics of leadership on interprofessional teams. *Acad Med.* 2012;87(12):1762-1767

Lingard L, Regehr G, Espin S, Whyte S. A theory-based instrument to evaluate team communication in the operating room: balancing measurement authenticity and reliability. *Qual Saf Health Care.* 2006;15(6):422-426.

Makary M, Mukherjee A, Sexton B et al. Operating room briefings and wrong site sugery. *J Am Coll Surg* 2007;204:236–243.

Mascherek AC, Schwappach DL, Bezzola P. Frequency of use and knowledge of the WHO-surgical checklist in Swiss hospitals: a cross-sectional online survey. *Patient safety in surgery.* 2013;7(1):36

Meinberg EG, Stern PJ. Incidence of wrong site surgery amongst hand surgeons. J Bone Joint Surg 2003; 85-A: 193-197.

Mitchell L, Flin R, Yule S, Mitchell J, Coutts K, Youngson G. Development of a behavioural marker system for scrub practitioners' non-technical skills (SPLINTS system). *J Eval Clin Pract.* 2013;19(2):317-323.

NASS. Sign, Mark, and Xray (SMaX): A Checklist for Safety. https://www.researchgate.net/figure/Sign-mark-and-X-ray-preoperative-check-list-NASS-Reprinted-from-Sign-mark-X-ray_fig1_236061354

North American Spine Society. Prevention of wrong-site surgery: sign, mark & x-ray (SMaX). LaGrange, IL: North American Spine Society, 2001.

Newman-Toker DE, Pronovost PJ. Diagnostic errors-the next frontier for patient safety. JAMA. 2009;301(10):1060–2. doi: 10.1001/jama.2009.249. [PubMed] [Cross Ref]

Nicolini D, Waring J, Mengis J. The challenges of undertaking root cause analysis in health care: a qualitative study. *J Health Serv Res Policy.* 2011;16 Suppl 1:34-41.

O'Connor P, Reddin C, O'Sullivan M, O'Duffy F, Keogh I. Surgical checklists: the human factor. Patient Saf Surg. 2013;7(1):14. [PMC free article] [PubMed]

Osbaldiston R. "It won't happen to me": The optimism bias. https://psychonline.eku.edu/insidelook/"it-won't-happen-me"-optimism-bias

Parker SH, Flin R, McKinley A, Yule S. Factors influencing surgeons' intraoperative leadership: video analysis of unanticipated events in the operating room. *World J Surg.* 2014;38(1):4-10.

Pikkel D, Sharabi-Nov A, Pikkel J. The importance of side marking in preventing surgical site errors. Int J Risk Saf Med. 2014;26(3):133–8. [PubMed]

Raemer DB, Kolbe M, Minehart RD, Rudolph JW, Pian-Smith MC. Improving Anesthesiologists' Ability to Speak Up in the Operating Room: A Randomized Controlled Experiment of a Simulation-Based Intervention and a Qualitative Analysis of Hurdles and Enablers. *Acad Med.* 2016;91(4):530-539.

Rasmussen J Skills, rules, and knowledge: signals, signs and symbols and other distinctions in human performance models. IEEE Transactions: Systems, Man & Cybernetics, 1983, SMC-13, pp.257-267. ISSN: 0018-9472.

Reason J. Human error: models and management. Br Med J. 2000;320:768–70. doi: 10.1136/bmj.320.7237.768. [PMC free article] [PubMed] [Cross Ref]

Reason, James "The Contribution of Latent Human Failures to the Breakdown of Complex Systems". Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences. 1990: 327 (1241): 475–484. doi:10.1098/rstb.1990.0090.

Rosenbaum, L. Cursed by Knowledge-Building a Culture of Psychological Safety. NEJM. 2019. 380;8:786-790.

Roussin CJ. Increasing trust, psychological safety, and team performance through dyadic leadership discovery. *Small Group Research.* 2008;39(2):224- 248.

Roussin CJ, MacLean TL, Rudolph JW. The safety in unsafe teams: A Multilevel Approach to Team Psychological Safety.*J Management* 2014:42: 1409-1433.https://doi.org/10.1177/0149206314525204

Rowlands J, Yeager MP, Beach M, Patel HM, Huysman BC, Loftus RW. Video observation to map hand contact and bacterial transmission in operating rooms. *Am J Infect Control.* 2014;42(7):698-701.

Rudolph JW, Simon R, Rivard P, Dufresne RL, Raemer DB. Debriefing with good judgment: combining rigorous feedback with genuine inquiry. *Anesthesiol Clin.* 2007;25(2):361-376.

Russ S, Rout S, Caris J, Mansell J, Davies R, Mayer E, et al. Measuring variation in use of the WHO surgical safety checklist in the operating room: a multicenter prospective cross-sectional study. J Am Coll Surg. 2015;220(1):1-11. e4. [PubMed]

Santiesteban L, Hutzler L, Bosco JA, 3rd, Robb W, 3rd. Wrong-Site Surgery in Orthopaedics: Prevalence, Risk Factors, and Strategies for Prevention. *JBJS Rev.* 2016;4(1). 1 doi: 10.2106/JBJS.RVW.O.00030

Sarker S, Vincent C. Errors in Surgery: Review. Intl J Surg. 2005: 75-81.

Sedgwick P, Greenwood N. Understand- ing the Hawthorne effect. BMJ. 2015 Sep 04;351:h4672. PMID: 26341898. http:// www.bmj.com/content/351/bmj.h4672. long.

Sendlhofer G, Lumenta DB, Leitgeb K, et al. The gap between individual perception and compliance: a qualitative follow-up study of the Surgical Safety Checklist Application. *PloS one.* 2016;11(2):e0149212.

Sewell M, Adebibe M, Jayakumar P, et al. Use of the WHO surgical safety checklist in trauma and orthopaedic patients. *International orthopaedics.* 2011;35(6):897-901

Singer SJ, Molina G, Li Z, et al. Relationship between operating room teamwork, contextual factors, and safety checklist performance. *Journal of the American College of Surgeons.* 2016;223(4):568-580. e562.

Sparks E, Wehbe-Janek H, Johnson R, et al. Surgical safety checklist compliance: a job done poorly! J Am Coll Surg 2013;217:867–73.

Stahel P. The 10th year of the "Universal Protocol": are our patients safer today? Bone Joint. 2014;360(3):7–10.

Taylor K, Mayell A, Vandenberg S, Blanchard N, Parshuram CS. Prevalence and indications for video recording in the health care setting in North American and British paediatric hospitals. Paediatr Child Health. 2011;16(7):e57-60.

The Lancet Commission on Global Surgery. 2015.
https://www.thelancet.com/commissions/global-surgery

WHO. https://www.who.int/patientsafety/safesurgery/checklist/en/

Wilson, James Q; Kelling, George L (Mar 1982), *"Broken Windows: The police and neighborhood safety"*, The Atlantic*, retrieved 2007-09-03 (Broken windows (PDF)*, Manhattan institute*))*

Wurster LA, Thakkar RK, Haley KJ, et al. Standardizing the initial resuscitation of the trauma patient with the Primary Assessment Completion Tool using video review. *J Trauma Acute Care Surg.* 2017;82(6):1002-1006.

Yule S, Gupta A, Gazarian D, et al. Construct and criterion validity testing of the NOTSS behavior assessment tool using videos of simulated operations. *Br J S.* 2018: https://onlinelibrary.wiley.com/doi/abs/10.1002/bjs.10779?%3Fcampaign%3DBJSurg_Twitter=%3Fcampaign%3DBJSurg_Twitter&hootPostID=00837941910c8cff05800ddf104777b5

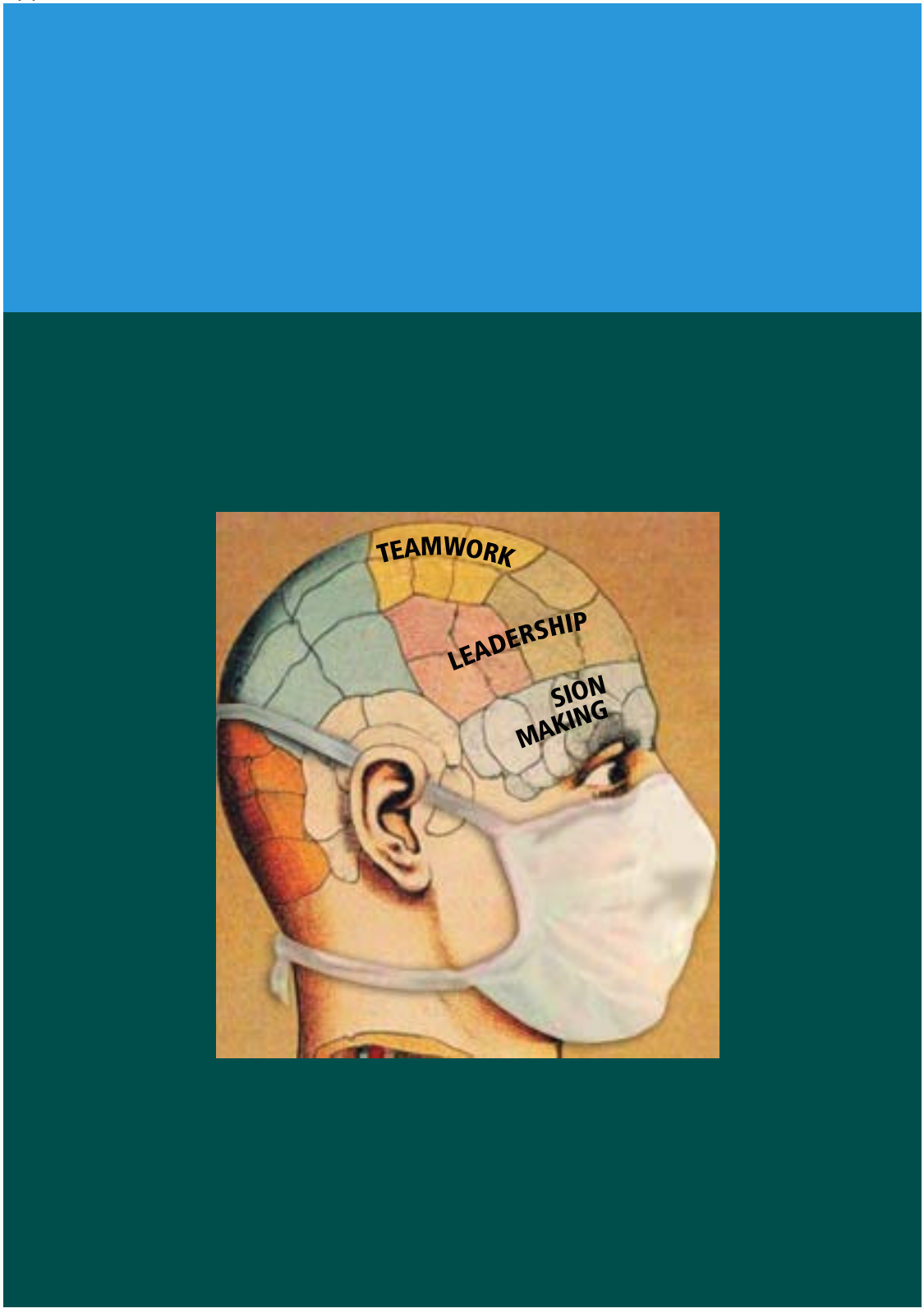
Yule S, Flin R, Maran N, Rowley D, Youngson G, Paterson-Brown S. Surgeons' non-technical skills in the operating room: reliability testing of the NOTSS behavior rating system. *World J Surg.* 2008;32(4):548-556.

Yule S, Flin R, Paterson-Brown S, Maran N. Non-technical skills for surgeons: a review of the literature. Surgery 2006;139:140–9.

Yule S, Smink D. Nontechnical skills in surgery. ACS Surgery: Principles and Practice. 2013: 1-8

Appendices

Appendix 1



Appendix 2

Appendix 3

PLINTS)

Structuring observation, rating and feedback of scrub practitioners' behaviours in the operating theatre

Appendix 4

BCH REAL EVENT ANALYSIS AND LEARNING (R.E.A.L.) OR COMPLIANCE CHECKLIST TOOL

DEMOGRAPHICS

GENERAL

Date (mo/dd/yyyy) _____

Location (hospital/OR#) _____

Procedure _____

Surgeon _____

CASE

CASE ID _____ **# IN OR SEQUENCE** _____

Start time (Pt entered OR) _____ **AM/PM**

Evaluator/Facilitator _____

Case Complexity _____

ITEM	SCORE *(circle best response)*		
1. Pre OR Huddles:	*These pre-OR discussions regarding the case, typically occur between 2 team member types prior to formal pre-op huddle. These can occur in room as well as hallways and pre-op holding.*		
a. Surgeon-Nurse	Observed	Not Observed	*Location:*
b. Surgeon-anesthesia	Observed	Not Observed	*Location:*
c. Nurse-anesthesia	Observed	Not Observed	*Location:*
2. Pre-Op Huddle:	*Occurs prior to patient arrival among team members within the OR.*		
a. Surgical plan	Done	Not Done	
b. Needed equipment	Done	Not Done	
c. Patient condition	Done	Not Done	
d. Associated medical conditions	Done	Not Done	
e. Anesthesia type and time	Done	Not Done	
3. Sign-in:	*Occurs between anesthesia and circulating nurse on arrival of patient in OR and prior to transfer to bed.*		
a. Patient id	Done	Not Done	
b. Procedure to be performed	Done	Not Done	
c. Surgical site marked	Done	Not Done	
d. Weight	Done	Not Done	
e. Known allergies	Done	Not Done	
f. Anesthesia plan	Done	Not Done	
4. Surgical start in OR	*First patient of the day, complete section a. All subsequent patients, complete section b.*		
a. First patient of the day: *Time of surgery start* _____	On-time *(≤30 mins from surgical start time)*	<15 delay	>15 delay
b. All subsequent patients: *Time of prior pt exit* _____	On-time *(≤30 mins from prior patient exit)*	<15 delay	>15 delay

BCH REAL EVENT ANALYSIS AND LEARNING (R.E.A.L.) OR COMPLIANCE CHECKLIST TOOL

5. Time Out:	*Occurs during formal time out prior to initiation of procedure.*			
a. Team introductions	Complete	Incomplete	Not Done	
b. Surgical procedure to be performed: **"Complete" = (1) mention of procedural details, (2) equipment, (3) post op casting and/or bracing*	Complete*	Incomplete	Not Done	
c. Surgical site ID	Done	Not Done		
d. Marking visible	Done	Not Done		
e. Consent matches	Checked	Not Checked		

6. Briefing – Surgeon		
a. Operative plan	Done	Not Done
b. Possible difficulties	Done	Not Done
c. Expected duration	Done	Not Done
d. Imaging if needed	Done	Not Done
e. Equipment needed	Done	Not Done
f. Fire safety	Done	Not Done
7. Briefing – Anesthesia		
a. Blood if needed	Done	Not Done
b. Anti-biotics given	Done	Not Done
8. Briefing - Team		
a. Allergies	Done	Not Done
b. Weight	Done	Not Done
c. Tourniquet pressure	Done	Not Done
d. Any safety concerns?	Done	Not Done
e. Speak up if issues arise	Done	Not Done

9. Intraoperative Essential Steps:	*Occur at various time points within the surgery. Note times.*			
a. Set up	Even-paced	Rushed	Slow	
b. Tourniquet time at 60 min	Done	Not Done	N/A	
c. Tourniquet time every 30 min after 60	Done	Not Done	N/A	
d. Fluoroscopy imaging review (if applicable)	Team-based/ collaborative	Single-provider	Not Reviewed	N/A
e. Equipment all present	No issues	Missing; Retrieved; Delay <5	Missing; Retrieved; Delay >5	Missing; Not Retrieved
f. Number of time circulator leaves and returns to room during case	Number of Times:			

10. Wound closure			
a. Wound inspection	Vocalized & Inspected	Vocalized only	Not Done
b. Correct count status: **"Complete" = Out loud readback/feedback Nurse and Circulator + informing surgeon of result + use of electronic wand over operative site*	Complete*	Incomplete	Not Done
11. Sign-out			
a. Name of procedure	Done		Not Done
b. Any concerns (e.g. patient, equipment)	Complete	Incomplete	Not Done
c. Anything we could have done better?	Complete	Incomplete	Not Done
d. TIME OUT OF ROOM	TIME:		

Comments:

E.g. evidence of scores; observations for team improvement; scoring concerns, etc.

Appendix 5A

REAL OR Tool to Assess Quality Characteristics of Debrief

11/8/18 version

Overarching Goals of Post Day Debrief:

- *Team Enculturation : Building Awareness of Value of Team-based Performance*
  - *Eg. Identify and Appreciate*
    - *Micro-climate Dynamics*
    - *Shared Mental Models*
- *Problem Identification : Overt and Latent Safety Threat Identification*
- *Problem Solving : Moving from individual to generalizable observations and change*
- *Provide Closure on the Day : Leaving the adult learner in a positive/productive place*

Overarching Quality Characteristics of Debriefing:

- *Overall Team **Engagement***
- ***Balanced**/Shared Discussion*
- ***Depth** of Discussion*
- ***Relevancy** of Discussion*
- ***Solution-oriented** Discussion*
- ***Tone** of Discussion*
- ***Pace/Tempo** of Discussion*

1. Please rate level of **Team Engagement** in discussion	1-5 Likert (to come)
a. To what extent were views **free flowing** vs. need for multiple prompts	
b. To what extent were views **self-derived** (originated from participant vs. suggested by the debriefer)	
c. To what extent was the whole team engaged (from 1-2 **predominant voices**, to 2-4, to >4 voices) throughout the discussion	
2. Please rate level of conversation **Balance/Shared** discussion	
a. To what extent was discussion **between participants** (vs. between participant and debriefer)	
b. To what extent did discussion involve **multidisciplinary** team members (vs. single speciality)	
3. Please rate **Depth** of discussion	
a. To what extent did discussion delve to more developed discussion (vs. remain on initial thoughts and ideas)?	
b. To what extent to one element of discussion build to another among the team (Eg. team “connects the dots”)	
4. Please rate **Relevancy** of discussion	

a. To what extent were topics relevant to the **day's activities**?	
c. To what extent did the following **topics inherent to Team behaviors** get raised and discussed: a. Sharing mental models: During the team's preparation/timeout/ongoing b. Microclimate dynamics: Subteam behaviors that limit communication/coordination	
d. To what extent did **personal stories** of related events play a role in the discussion?	
e. <u>Please list problems/challenges/topics here: *(open answer format)*</u>	
5. Please rate how **Solution-oriented** the discussion was	
a. To what extent did the team begin to **ideate solutions** to identified problems or challenges (vs. left conversation with only statements of problems)?	
b. To what extent were solutions **developed collaboratively** via build from one idea to the next (vs. from single participant)?	
c. To what extent did **solutions emanate from participant** ideas (vs. debriefer prompts/suggestions)?	
d. <u>Please list solutions here: *(open answer format)*</u>	
6. Please describe the **Tone** of the discussion	
a. **Debriefer**: (Disarming, welcoming, engaging vs. Accusatory, Evaluative, Etc)	
b. **Participants**: (Up, Enjoyable, Light vs. "Under the light")	
c. Between Participants: Agreement vs. Dissonance	
7. Please describe the **Tempo** of the discussion (Total time of Debrief:_____)	
a. **Organized** discussion (vs. unclear facilitation)	
b. Comfortable **flow** (vs. felt rushed)	
c. **Conversational** (moved nicely from topic to topic vs. Didactic)	

Appendix 5B
January 16, 2019 version

R.E.A.L. OR Tool to Assess Quality Characteristics of Debrief

Overarching Goals of Post Day Debrief:

- *Team Enculturation : Building Awareness of Value of Team-based Performance*
  - *Suggested topics may include:*
    - *Identify and Appreciate*
    - *Micro-climate Dynamics*
    - *Shared Mental Models*
- *Problem Identification : Overt and Latent Safety Threat Identification*
- *Problem Solving : Moving from individual to generalizable observations and change*
- *Provide Closure on the Day : Leaving the adult learner in a positive/productive place*

Overarching Quality Characteristics of Debriefing:

- *Overall Team **Engagement***
- ***Balanced**/Shared Discussion*
- ***Depth** of Discussion*
- ***Relevancy** of Discussion*
- ***Solution-oriented** Discussion*
- ***Tone** of Discussion*
- ***Pace/Tempo** of Discussion*

DEMOGRAPHICS

DATE:

FACILITY

SURGEON:

FACILITATOR:

ATTENDEES:

SCORER:

8.	Please rate level of **Team Engagement** in discussion					
	d. To what extent were views **free flowing** vs. need for multiple prompts	Never[1]	rarely	Sometimes	often	All the time
	e. To what extent were views **self-derived** (originated from participant vs. suggested by the debriefer)	Never	rarely	Sometimes	often	All the time
	f. To what extent was the whole team engaged. Number of **predominant voices**: 0, 1, 2, 3-4, >4) throughout the discussion	0	1	2	3-4	>4
9.	Please rate level of conversation **Balance/Shared** discussion					
	f. To what extent was discussion **between participants** (vs. between participant and debriefer)	Never	rarely	Sometimes	often	All the time
	g. To what extent did discussion involve **multidisciplinary** team members (vs. single specialiy)	Never	rarely	Sometimes	often	All the time
10.	Please rate **Depth** of discussion					
	c. To what extent did discussion delve to **more developed discussion** on topic (vs. remain on initial thoughts and ideas)?	Never	rarely	Sometimes	often	All the time
	d. To what extent did one element of **discussion diverge** to another among the team (Eg. team "connects the dots")	Never	rarely	Sometimes	often	All the time
11.	Please rate **Relevancy** of discussion					
	a. To what extent were topics relevant to the **day's activities**?	Never	rarely	Sometimes	often	All the time
	b. To what extent did the following **topics inherent to Team behaviors** get raised and discussed: Eg. a. **Sharing mental models**: During the team's preparation/timeout/ongoing b. **Microclimate dynamics**:	Never	rarely	Sometimes	often	All the time

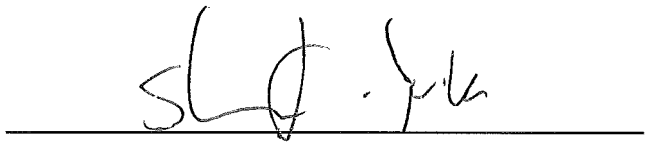
[1]Frequency – 5 point, Vagias, Wade M. (2006). *"Likert-type scale response anchors.* Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management. Clemson University

	Subteam behaviors that limit communication/coordination					
c.	To what extent did **personal stories** of related events play a role in the discussion?	Never	rarely	Sometimes	often	All the time
d.	**<u>Please list problems/challenges/topics/notable successes here:</u>** ***<u>(MANDATORY; open answer format)</u>***					

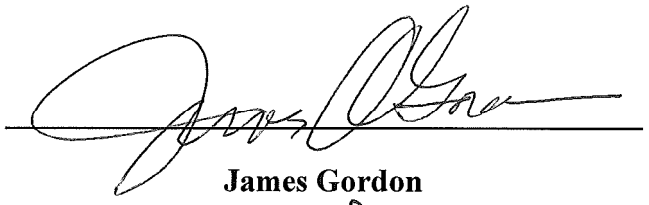
12.	Please rate how **Solution-oriented** the discussion was					
	e. To what extent did the team begin to **ideate solutions** to identified problems or challenges (vs. left conversation with only statements of problems)?	Never	rarely	Sometimes	often	All the time
	f. To what extent were solutions **developed collaboratively** via build from one idea to the next (vs. from single participant)?	Never	rarely	Sometimes	often	All the time
	g. To what extent did **solutions emanate from participant** ideas (vs. debriefer prompts/suggestions)?	Never	rarely	Sometimes	often	All the time
	h. <u>Please list solutions to problems/reinforcements of successes here: (*MANDATORY, open answer format)*</u>					

13. Please describe the **Tone** of the discussion					
d. **Debriefer**: (**Disarming, welcoming, engaging** vs. Accusatory, Evaluative, Etc)	Never	rarely	Sometimes	often	All the time
e. **Participants**: (**Up, Enjoyable, Light** vs. "Under the light")	Never	rarely	Sometimes	often	All the time
f. Between Participants: **Agreement** vs. Dissonance	Never	rarely	Sometimes	often	All the time
14. Please describe the **Tempo** of the discussion (Total time of Debrief:_____)					
d. **Organized** discussion (vs. unclear facilitation)	Never	rarely	Sometimes	often	All the time
e. Comfortable **flow** (vs. felt rushed)	Never	rarely	Sometimes	often	All the time
f. **Conversational** (moved nicely from topic to topic vs. Didactic)	Never	rarely	Sometimes	often	All the time

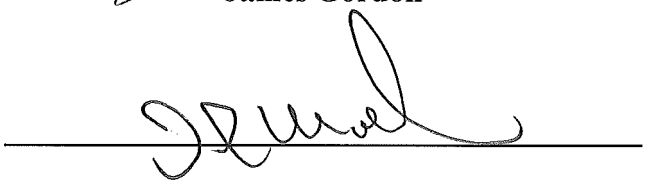
This Thesis, Real Event Learning and Analysis (REAL): Assessing and Improving Surgical Team Performance, presented by Peter M. Waters, and Submitted to the Faculty of The Harvard Medical School in Partial Fulfillment of the Requirements for the Master of Medical Sciences in Medical Education has been read and approved by:



Steven J. Yule



James Gordon



J. Lawrence Marsh

Date: April 29, 2019