Nontechnical Skills in Postgraduate Surgical Education

by

Nicolas Joachim Dedy

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Institute of Medical Science
University of Toronto

© Copyright by Nicolas Joachim Dedy 2015
Nontechnical Skills in Postgraduate Surgical Education

Nicolas Joachim Dedy

Doctor of Philosophy
Institute of Medical Science
University of Toronto
2015

Abstract

**Background**: Surgeons’ nontechnical skills are critical for patient safety in the operating room (OR) and have been included in resident training requirements. The optimal approach to their training and assessment though is unknown. The goal of this thesis was to develop a structured approach to the training and assessment of nontechnical skills in surgical education.

**Methods**: Objectives were: (1) to establish the prevalence of nontechnical training components in postgraduate curricula; (2) to develop a reliable tool for the criterion-referenced assessment of residents’ nontechnical performance; (3) to determine surgical residents’ innate nontechnical skills and how these relate to personality; (4) to evaluate the effectiveness of a strategy for competency-based training of nontechnical skills in surgical residency.
**Results:** (1) One-third of responding residency programs reported nontechnical training components; (2) the Objective Structured Assessment of Nontechnical Skills (OSANTS) tool was developed to assess resident-specific skills. Good psychometric properties were demonstrated, both in simulation and in the real OR; (3) large inter-individual differences in nontechnical performance were found among 31 junior surgical residents. Some variability could be explained by personality traits; (4a) a basic skills curriculum was evaluated in a randomized trial. Curriculum trained first-year residents (n=11) outperformed conventionally trained peers (n=11) in knowledge and attitudes about nontechnical skills. Curriculum-trained residents’ performance in a simulated OR improved significantly from baseline to post-training (OSANTS: (median 23(interquartile range 17-28) vs. 31(25-33), p=0.012), while conventionally trained residents did not improve (OSANTS, 26(24-32) vs. 24(23-32), p=0.713); (4b) In an interrupted time-series study, observed nontechnical performance of senior residents (n=11) in the OR improved significantly following one debriefing and feedback session.

**Conclusion:** Based on the results, a combined approach to nontechnical education in surgical postgraduate education is proposed, with a basic skills curriculum in the first year, and debriefing and feedback in formative assessments throughout residency training.
Acknowledgments

This thesis would not have been possible without the help of all the fantastic people I was fortunate to work with during the past three and a half years, who supported me, helped me with my research, and offered their friendship.

First, I would like to express my sincerest gratitude to my supervisor Dr. Teodor Grantcharov who guided and supported me through this project, was always available for help and advice, but also taught me to work independently, preparing me for the role as an independent researcher. Thank you, Teodor, for giving me the unique opportunity to join your research group and pursue this degree, for your continuous support, and for your friendship.

I would like to extend my gratitude to the members of my supervisory committee, Dr. Carol-Anne Moulton and Dr. Najma Ahmed, for their invaluable advice and input, constructive criticism and feedback, and support throughout my graduate studies.

I wish to thank Dr. Boris Zevin who introduced me to the lab, contributed to several research projects, motivated and supported me with his optimism, and most importantly has become a close and trusted friend.

I further thank my fellow researchers Dr. Andras Fecso, Dr. Heinrich Husslein, Dr. Marisa Louridas, Dr. Peter Szasz, and Dr. Eli Shore for countless discussions about research, for their constructive feedback, help with my studies, and for their friendship.

Moreover, I wish to thank Dr. Jennifer Hickey and Karthik Raj for the help with my research, and the continued support in practical and administrative questions.

I wish to express my appreciation of the staff of the Allan Waters Family Simulation Centre, including Roger Chow, Dr. Niranjan Sathivel, Ashley Rosen, and Sue Zelko, for their continued support of my research. Especially Roger Chow’s enthusiasm, creative ideas and experience in simulation were vital for the success of my randomized trial.
I further wish to acknowledge all the residents at the University of Toronto, Department of Surgery, for taking the time to participate in my studies, as well as the staff surgeons at St. Michael’s Hospital for allowing me to observe residents during their lists in the operating room.

Finally, I owe my deepest gratitude to my wife Dr. Esther Bonrath who has shared all successes and setbacks and has supported me through good and bad times with knowledge, advice, and with her love. Thank you for everything.
Contributions

I, Nicolas Joachim Dedy, solely prepared this thesis and am the first author of the six manuscripts directly resulting from this thesis research.

As such, I designed all studies and performed all aspects of this work including the planning, execution, data analysis, and writing of the manuscripts. I formally acknowledge the following contributions of other individuals:

Dr. Teodor P. Grantcharov (Supervisor and Program Advisory Committee member): mentorship; laboratory resources; guidance and assistance in planning, execution, and data analysis; critical review of all manuscripts and thesis.

Dr. Najma Ahmed (Program Advisory Committee member): mentorship; guidance and assistance in planning of studies; assistance in execution of the study in chapter six; critical review of the manuscript for chapter six and the thesis.

Dr. Carol-Anne Moulton (Program Advisory Committee member): mentorship; guidance and assistance in planning of studies, and critical review of the thesis.

Dr. Esther M. Bonrath: assistance in planning, execution, data analysis, and critical review of all manuscripts; data collection for the study detailed in chapters four, five and six.

Dr. Boris Zevin: assistance in planning, execution, data analysis and critical review of the manuscripts of the systematic review detailed in chapter one and the study detailed in chapter three.

Dr. Peter Szasz: assistance in execution, data collection and analysis, and critical review of the manuscript of the studies detailed in chapters four and chapter seven.

Dr. Heinrich Husslein: assistance in execution, data analysis and critical review of the manuscript of the study detailed in chapters four and five.

Dr. Marisa Louridas: assistance in execution, data collection and analysis, and critical review of the manuscript of the study detailed in chapter four.
Dr. Andras B. Fecso: assistance in execution, data collection and analysis, and critical review of the manuscript of the study detailed in chapter seven.

Professor R. Michael Bagby: assistance in planning, data analysis and critical review of the manuscript of the study detailed in chapter five.

Shauna C. Kushner: assistance in data analysis and critical review of the manuscript of the study detailed in chapter five.

Corinne Holubowich (medical librarian): assistance with the systematic literature review reported in chapter one.
# Table of Contents

Acknowledgments .................................................................................................................. iv
Contributions .......................................................................................................................... vi
Table of Contents ..................................................................................................................... viii
Abbreviations ........................................................................................................................ xiii
List of Tables ............................................................................................................................ xvi
List of Figures .......................................................................................................................... xviii
List of Appendices .................................................................................................................... xix

1 GENERAL INTRODUCTION ................................................................................................. 1

1.1 History of Nontechnical Skills ......................................................................................... 2

1.1.1 Human Factors Research in Aviation ........................................................................... 2
1.1.2 Adverse Events and Human Error in Operative Patient Care ......................................... 9

1.2 Nontechnical Skills in Surgery ..................................................................................... 17

1.2.1 Terminology and Definition ....................................................................................... 17
1.2.1.1 Definition and Characterization of Nontechnical Skills ........................................ 18
1.2.1.1.1 Cognitive skills ........................................................................................................ 20
1.2.1.1.2 Social skills ............................................................................................................... 32
1.2.1.1.3 Personal behaviours ............................................................................................... 39
1.2.2 Training Interventions and Initiatives ......................................................................... 40
1.2.2.1 CRM- and team-training interventions ..................................................................... 41
1.2.2.2 Simulation-based training .......................................................................................... 46
1.2.3 Nontechnical Skills in Postgraduate Surgical Education ............................................... 49
1.2.3.1 How do surgical residents learn nontechnical skills? ................................................ 49
1.2.3.2 Factors influencing Innate Nontechnical Skills and Skill Development in Residency Training ................................................................................................................. 54
3 CURRENT CONCEPTS OF TEAM TRAINING IN SURGICAL RESIDENCY: A SURVEY OF NORTH AMERICAN RESIDENCY PROGRAM DIRECTORS

3.1 Abstract
3.2 Introduction
3.3 Methods
3.4 Results
3.5 Discussion
3.6 Conclusion

4 OBJECTIVE STRUCTURED ASSESSMENT OF NONTECHNICAL SKILLS: RELIABILITY OF A GLOBAL RATING SCALE FOR THE IN-TRAINING ASSESSMENT IN THE OPERATING ROOM

4.1 Abstract
4.2 Introduction
4.3 Methods
4.4 Results
4.5 Discussion

5 SURGEONS’ PERSONALITY CAN INFLUENCE THEIR NONTECHNICAL SKILLS IN AN INTRAOPERATIVE CRISIS

5.1 Abstract
5.2 Introduction
5.3 Methods
5.4 Results
5.5 Discussion

6 STRUCTURED TRAINING TO IMPROVE NONTECHNICAL PERFORMANCE OF JUNIOR SURGICAL RESIDENTS IN THE OPERATING ROOM: A RANDOMIZED CONTROLLED TRIAL

6.1 Abstract
6.2 Introduction
6.3 Methods ........................................................................................................................................165
6.4 Results........................................................................................................................................171
6.5 Discussion....................................................................................................................................175
6.6 Conclusion...................................................................................................................................180

7 IMPLEMENTATION OF AN EFFECTIVE STRATEGY FOR TEACHING NONTECHNICAL SKILLS IN THE OPERATING ROOM: A SINGLE-BLINDED NON-RANDOMIZED TRIAL ..........................................................181
7.1 Abstract.......................................................................................................................................181
7.2 Introduction...................................................................................................................................182
7.3 Methods .......................................................................................................................................184
7.4 Results..........................................................................................................................................187
7.5 Discussion.....................................................................................................................................190
7.6 Conclusion...................................................................................................................................195

8 GENERAL DISCUSSION ...................................................................................................................196
8.1 Summary of Study Findings ........................................................................................................196
8.2 Structuring nontechnical education in surgical residency training .............................................199
8.3 Implementation of a basic nontechnical skills curriculum ..........................................................202
8.4 Factors influencing the Development of Nontechnical Skills ....................................................205
8.5 Assessment of nontechnical skills ...............................................................................................207
8.6 Summary.....................................................................................................................................212

9 LIMITATIONS ................................................................................................................................214
9.1 Study context and participants .....................................................................................................214
9.2 Complex interventions ................................................................................................................215
9.3 Long term effects and patient outcome ........................................................................................216
9.4 Relationship between personality and nontechnical skills .........................................................218
9.5 Influence of the Hidden Curriculum ............................................................................................221
10 FUTURE DIRECTIONS .................................................................................................................. 222

10.1 Formative Assessments by Staff Surgeons ........................................................................ 222

10.2 Psychometric Properties of OSANTS when used by Staff Surgeons .......................... 222

10.3 Implications of Personality Testing for Surgical Education .............................................. 223

10.4 Inducing a Culture Change in the Educational Environment ........................................... 224

References ................................................................................................................................... 226

Appendices ................................................................................................................................ 256
# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABMS</td>
<td>American Board of Medical Specialties</td>
</tr>
<tr>
<td>ABS</td>
<td>American Board of Surgeons</td>
</tr>
<tr>
<td>ACGME</td>
<td>Accreditation Council for Graduate Medical Education</td>
</tr>
<tr>
<td>ACS</td>
<td>American College of Surgeons</td>
</tr>
<tr>
<td>AERA</td>
<td>American Educational Research Association</td>
</tr>
<tr>
<td>ANTS</td>
<td>Anaesthesiologists' Non-Technical Skills</td>
</tr>
<tr>
<td>APA</td>
<td>American Psychological Association</td>
</tr>
<tr>
<td>APDS</td>
<td>Association of Program Directors in Surgery</td>
</tr>
<tr>
<td>BL</td>
<td>Baseline</td>
</tr>
<tr>
<td>C-HRV</td>
<td>Coefficient of Heart Rate Variability</td>
</tr>
<tr>
<td>CRM</td>
<td>Cockpit-, Crew-, or Crisis Resource Management</td>
</tr>
<tr>
<td>DoD</td>
<td>Department of Defense</td>
</tr>
<tr>
<td>FAA</td>
<td>Federal Aviation Administration</td>
</tr>
<tr>
<td>FFM</td>
<td>Five-Factor Model of Human Personality</td>
</tr>
<tr>
<td>GRADE</td>
<td>Grading of Recommendations, Assessment, Development and Evaluation</td>
</tr>
<tr>
<td>ICC</td>
<td>Intra-class Correlation Coefficient</td>
</tr>
<tr>
<td>ITER</td>
<td>In-Training Evaluation Report</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>JAA</td>
<td>Joint Aviation Authorities</td>
</tr>
<tr>
<td>LOFT</td>
<td>Line Oriented Flight Training</td>
</tr>
<tr>
<td>MTT</td>
<td>Medical Team Training</td>
</tr>
<tr>
<td>NASA</td>
<td>National Aeronautics and Space Administration</td>
</tr>
<tr>
<td>n.d.</td>
<td>No date</td>
</tr>
<tr>
<td>NEO-FFI</td>
<td>Neuroticism-Extraversion-Openness Five-Factor Inventory</td>
</tr>
<tr>
<td>NEO-PI-R</td>
<td>Neuroticism-Extraversion-Openness Personality Inventory Revised Version</td>
</tr>
<tr>
<td>NOTECHS</td>
<td>Non-Technical Skills</td>
</tr>
<tr>
<td>NOTSS</td>
<td>Non-Technical Skills for Surgeons</td>
</tr>
<tr>
<td>OCHRA</td>
<td>Observational Clinical Human Reliability Assessment</td>
</tr>
<tr>
<td>OED</td>
<td>Oxford English Dictionary</td>
</tr>
<tr>
<td>OR</td>
<td>Operating Room</td>
</tr>
<tr>
<td>ORMAQ</td>
<td>Operating Room Management Attitudes Questionnaire</td>
</tr>
<tr>
<td>OSANTS</td>
<td>Objective Structured Assessment of Nontechnical Skills</td>
</tr>
<tr>
<td>OSATS</td>
<td>Objective Structured Assessment of Technical Skills</td>
</tr>
<tr>
<td>OTAS</td>
<td>Observational Teamwork Assessment for Surgery</td>
</tr>
<tr>
<td>PBA</td>
<td>Procedure-Based Assessment</td>
</tr>
<tr>
<td>PD</td>
<td>Program Director</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PT</td>
<td>Post-training</td>
</tr>
<tr>
<td>SAQ</td>
<td>Safety Attitudes Questionnaire</td>
</tr>
<tr>
<td>SBAR</td>
<td>Situation-Background-Assessment-Recommendation</td>
</tr>
<tr>
<td>TeamSTEPPSTM</td>
<td>Team Strategies and Tools to Enhance Performance and Patient Safety</td>
</tr>
<tr>
<td>T-TAQ</td>
<td>TeamSTEPPSTM Team Attitudes Questionnaire</td>
</tr>
<tr>
<td>U.K.</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>U.S.</td>
<td>United States</td>
</tr>
<tr>
<td>VA</td>
<td>Department of Veterans Affairs</td>
</tr>
</tbody>
</table>
List of Tables

Table 1: Strength of body of evidence, organized by main outcome themes .......................... 71

Table 2: Non-Technical Skills for Surgeons (NOTSS) rating framework, version 1.2. ..... 102

Table 3: Validity evidence supporting construct validity of scores generated with the NOTSS system ............................................................................................................. 104

Table 4: Categories and elements of the Oxford NOTECHS system .................................... 106

Table 5: Validity evidence supporting scores generated with the Oxford NOTECHS rating system .................................................................................................................. 108

Table 6: OTAS framework for the surgical team during the intraoperative phase. ............. 110

Table 7: Validity evidence supporting scores generated with the OTAS rating system. .... 112

Table 8: Results from raters 1 and 2 scoring surgical trainees in a simulated environment (recorded crisis simulations) ........................................................................................................ 138

Table 9: Results from both raters scoring surgical trainees during live-observations in the OR ......................................................................................................................... 139

Table 10: Agreement between two raters scoring 31 surgical trainees in a simulated environment (recorded crisis simulations) ........................................................................................................ 140

Table 11: Trainees’ scores on the NEO-FFI ........................................................................... 155

Table 12: Aggregate scores from both raters’ assessment of trainees’ nontechnical performance ................................................................................................................... 156

Table 13: Nontechnical Skills Training Curriculum as administered to the intervention group ......................................................................................................................................... 168

Table 14: Participant demographics at baseline .................................................................... 172
Table 15: Scores of conventional and curriculum-trained groups at baseline and post-training. .................................................................................................................................................................................... 173

Table 16: Free-text comments of participants after receiving the non-technical skills curriculum-training ................................................................................................................................................................................................................................................... 175

Table 17: Deficiencies in nontechnical performance observed by surgeon-educators and addressed in the debriefing and feedback discussions with residents. ................................................. 189
List of Figures

Figure 1: The dynamics of accident causation ................................................................. 6

Figure 2: Endsley's model of situation awareness in dynamic systems.......................... 21

Figure 3: Aviation decision process model. ..................................................................... 28

Figure 4: Experiential Learning Cycle. ............................................................................. 52

Figure 5: Search and exclusion algorithm in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses format ............................................................................ 70

Figure 6: Basic framework of rating systems for the assessment of performance based on observable behaviours in high-risk work environments................................................. 98

Figure 7: Methods currently used for CRM and team training among responding programs. ......................................................................................................................................................... 124

Figure 8: Training methods recommended by responding PD for resident CRM curriculum. ......................................................................................................................................................... 125

Figure 9: CONSORT diagram illustrating participant flow throughout the study period... 166
List of Appendices

Appendix 1: Overview of studies regarding target population, design, risk of bias, and comments including potential limitations...............................................................246

Appendix 2: Summary of study data pertaining to instructional design, learning objectives/ content, methods used for assessment, and outcome.................................249

Appendix 3: Main outcome themes and factors considered to upgrade/ downgrade strength of evidence.........................................................................................260

Appendix 4: Objective Structured Assessment of Non-Technical Skills (OSANTS Global Rating Scale)........................................................................................................263
1
GENERAL INTRODUCTION

In the first chapter I will outline the history of nontechnical skills, the evolution of training and assessment strategies of these skills in high reliability industries such as aviation, and their relevance in healthcare and specifically in surgery. I will further discuss the significance of non-technical skills in postgraduate surgical education, and review factors that may determine innate nontechnical skills and the development of these important competencies during surgical residency training. Finally, I am going to provide an overview of strategies to teach and assess nontechnical aspects of surgeons’ performance.

It must be clarified at this point that my focus in the present work is on nontechnical skills as they relate to the operating room (OR). Although nontechnical skills are also required in non-operative areas of the surgical profession including ward duties, outpatient clinics, and the emergency room, the OR represents the center of activity of a surgeon. Moreover, the OR is one of the most complex and demanding work environments within health care, requiring the surgeon to perform under stressful and sometimes unpredictable conditions with the ubiquitous risk of inadvertent patient harm. In a summary of the existing knowledge base of nontechnical skills in surgical postgraduate education I will then direct the reader’s attention to knowledge gaps and open questions that form the basis of this thesis.

The literature review for the general introduction was conducted using the databases Ovid MEDLINE (including publications from 1946 through November 2014), PsycINFO (1806 through December 2014), Embase Classic (1947 through 1979), Embase (1980 through 2014), and the Cochrane Library. Consensus papers, online databases of regulatory authorities, and meeting reports were identified through searches using www.google.ca and www.scholar.google.ca. In addition, hand searches for author names were conducted using cross-references. If applicable, books were accessed online or retrieved as hard copies through the University of Toronto Gerstein Science Information Centre.
1.1 History of Nontechnical Skills

1.1.1 Human Factors Research in Aviation

Early reports on human factors research in aviation date back as far as 1947, when Paul Fitts and Richard Jones were commissioned by the United States Air Force to explore design features of aircraft instruments “to improve pilot efficiency and reduce the frequency of accidents…” (p.1) (Fitts & Jones, 1947). Fitts and Jones (1947) subsequently analyzed 270 accidents involving “pilot error”, making recommendations to avoid similar adverse events. Although their report was predominantly focused on errors related to human-instrument interactions, many of the errors identified and described by the authors did in fact represent important cognitive, interpersonal, and personal resource skills that would four decades later be described as crew resource management, or nontechnical skills, respectively. Examples include errors resulting from a failure to understand or interpret situational information from the environment, described as “Errors in interpreting the Altimeter and Other Multi-Resolution Instruments” (p.16) (Fitts & Jones, 1947); errors due to “…’personal’ factors such as division of attention, excitement and fatigue…” (p.43) (Fitts & Jones, 1947); errors pertaining to communication failures including misinterpretation of hand signals, warning sounds and radio signals; errors pertaining to pilot leadership, described as “confusion regarding which pilot has the controls (in aircraft with tandem seating arrangements)” (p.29) (Fitts & Jones, 1947). Interestingly, in addition to recommendations regarding instrument design, Fitts and Jones (1947) advocated for standard operating procedures and the training of pilots: to avoid “forgetting errors” (p.44), they recommended the development of a mechanical checklist to be completed before take-off and landing, a concept that was later introduced in all routine flights of commercial and military aviation (Fitts & Jones, 1947). To address the problem of miscommunications and ambiguity regarding pilot leadership, they suggested “That the policy and training program with respect to the use of hand signals be examined with a view to complete or almost complete elimination of this method of communication and the substitution when necessary of mechanical signaling devices or spoken command” (Fitts & Jones, 1947). In the subsequent 30 years, as aircraft became more reliable and crashes occurred less frequently, research
focused increasingly on the human factor in aviation, and the role of human error in accidents (Helmreich & Foushee, 2010). Citing data from the Boeing Aircraft Company, Helmreich and Foushee (2010) pointed out that, in the time period from 1959 through 1979, more than 70 percent of severe accidents in commercial aviation worldwide could be attributed primarily to human factors. In the early 1970s, the National Aeronautics and Space Administration (NASA) began a research program at the Ames Research Center in Moffett Field, California to study the role of human factors in aviation safety (G. Cooper, White, & Lauber, 1980). In confidential interviews with airline crew members, issues pertaining to the training of new flight captains were most frequently addressed by interviewees, including questions of effective leadership, crew coordination, and communication (G. Cooper et al., 1980). In one of the first simulation studies, Ruffel Smith (1979) evaluated flight captains during full-mission flight simulations and observed increased error rates during crisis situations with higher workload (Ruffell Smith, 1979). Moreover, he identified a relationship between errors and vigilance, decision making, and management of available resources (Ruffell Smith, 1979). In the third component of the research program, NASA experts analyzed data from 62 accidents of commercial aircraft between 1968 and 1976, identifying problematic behaviours that contributed to the accidents. An example was a fully functional aircraft that crashed into the Florida Everglades in December 1972 killing all passengers and crew members on board, because all three crew members on the flight deck were preoccupied with a faulty indicator lamp and failed to notice the inadvertent deactivation of the autopilot and the subsequent gradual loss of altitude (G. Cooper et al., 1980). According to the authors, the most frequently observed human errors and behaviours contributing to the analyzed accidents were preoccupation with minor problems, deficiencies in leadership, failure to delegate tasks and set priorities, inadequate monitoring, failure to use all available data, and failure to communicate plans (G. Cooper et al., 1980). The researchers classified all observed problems into five main categories: social and communication skills (e.g., assertiveness, nonverification of communication, and assumptions); leadership and management skills (e.g., delegating, crew coordination, and prioritizing); planning, problem solving, and decision skills (e.g., problem solving strategies, decision under stress, and group think); role clarity (e.g., command responsibility, responsibility of first officer when captain deviates from safe practices); and
resource management (e.g., material, equipment, textual information, environmental information). The identified safety-relevant behaviours, skills and attitudes were subsumed under the concept of Resource Management in the cockpit, which marked the beginning of Cockpit Resource Management (CRM) training. John Lauber, a psychologist who was part of the NASA research team to first propose the training of resource management skills, defined CRM as “using all available resources – information, equipment, and people – to achieve safe and efficient flight operations” (p.20) (Lauber, 1984). After implementation of the first CRM training program by United Airlines in 1981, CRM training courses were subsequently introduced by most major airlines across North America, as well as airlines around the world (Helmreich, Merritt, & Wilhelm, 1999). Frequently, CRM courses were augmented with full-mission flight simulations, known as Line Oriented Flight Training (LOFT). Although the initial rationale behind CRM training was the reduction of errors, the first programs were predominantly focused on aspects of leadership, specifically targeting assertiveness of junior crew members to challenge decisions of the captain, and conversely flight captains’ acceptance of input from subordinate team members (Helmreich et al., 1999). The CRM concept of flat hierarchies on the flight deck, among other concepts, was at first not well-received in many countries with cultures that were strongly based on hierarchies, including many Asian and South American countries (Helmreich et al., 1999). National and organizational culture was subsequently identified as an important factor to be accounted for in the development and implementation of CRM programs. In the mid-1980s, CRM courses became more team oriented, including concepts such as team building and briefing, which was symbolically emphasized by a name change from Cockpit Resource Management to Crew Resource Management (Helmreich et al., 1999). Increasingly, CRM programs began to include concepts such as situation awareness, decision making, stress management, and error mitigation in the courses, which were now also administered to flight attendants, maintenance personnel and other crewmembers outside the flight deck (Helmreich et al., 1999). In 1990, the Federal Aviation Administration (FAA) in the United States (U.S.) made it mandatory for all major airlines to offer CRM training and LOFT to all crewmembers. Following the pioneering work in the U.S. all major airlines worldwide subsequently implemented components of CRM training in their pilot and flight crew training programs. By the mid-1990s, the content of CRM training had made a transition
towards concepts and strategies to avoid, mitigate and trap error, returning to the original purpose that had been proposed in 1979 on the NASA workshop (Helmreich et al., 1999). The fifth generation of CRM training was based in part on the work of Professor James Reason, who had developed a theory on human error and adverse events in complex systems (Reason, 1990). Reason’s theory was based on the observation that errors will always happen when humans are involved, even in the most reliable organizations with highly trained specialists (Reason, 1990). Reason differentiated between active and latent failures, with active failure being unsafe acts, errors, and mistakes enacted by individuals, and latent failures representing wrong decisions and unsafe conditions at the managerial level (e.g., understaffing, long work hours with sleep deprivation) (Reason, 1990). While active failures have immediate effects, latent failures can exist for a long time before leading to an adverse event (Reason, 1995). Most of the time, adverse events are the result of several different failures (active and latent), often in combination with specific unsafe conditions or other triggering effects (Reason, 1995). Reason (1990) views modern high-reliability organizations such as aviation or healthcare as complex systems with multiple layers of defence protecting against adverse events (Figure 1). Active and latent failures represent potential dangers that may lead to adverse events but are stopped by the system’s defences (e.g., monitoring equipment, alarms, etc.). Certain local conditions and triggers, however, can create holes in the defences allowing for failures to go unnoticed and adverse events to occur (Reason, 1990). In order to prevent similar events in the future, it is critical to identify all factors, active and latent failures as well as local triggers and conditions, leading up to the event and not only focus on the failure that immediately caused the event (Reason, 1995). In the system approach to error, as advocated by Reason (1990) and implemented in CRM programs in aviation, all potential causes and contributing factors are considered, both at the “front line”, and at the organizational level, and defences are enhanced to prevent future adverse events (Reason, 1995). CRM training of the fifth generation was aimed at enhancing teamwork skills and behaviours of flight crews in order to avoid, trap and mitigate error and ultimately prevent adverse outcome (Helmreich et al., 1999).
Figure 1: The dynamics of accident causation

Active failures (unsafe acts) and latent failures represent potential dangers. Several layers of defence protect against these dangers. Holes in the defences represent “windows of opportunity” for an accident, which may occur if the holes line up, i.e. all defences fail. Holes in defence may be caused by local triggers, atypical conditions, and intrinsic defects of the system.


With permission from Cambridge University Press. © 1990, Cambridge University.

In Europe, the Joint Aviation Authorities (JAA) made the training of CRM skills, which were referred to by the European authorities as “nontechnical skills”, mandatory for all major airlines. Consequently, the JAA commissioned the NOTECHS project, a collaborative
project of national aviation authorities in Germany, France, and the Netherlands, as well as the human factors research group at the University of Aberdeen, Scotland, to develop a framework for the training and assessment of CRM or nontechnical skills of flight crews (van Avermaete, 1998). In their project report, the NOTECHS group proposed a framework of nontechnical skills of flight crews, comprising four main skill categories: Cooperation; Leadership and Managerial Skills; Situation Awareness; and Decision Making (Flin et al., 2003). European aviation authorities subsequently referred to the skills targeted by CRM training as nontechnical skills, but continued using the term CRM training when describing the teaching of nontechnical skills (van Avermaete, 1998).

The effectiveness of CRM training has been the subject of multiple studies since its first implementation, and CRM programs have been evaluated on various outcome parameters, including perceived usefulness, changes in attitudes towards CRM skills, changes in observed behaviour, and reduction in accidents and incidents. Helmreich and Wilhelm (1991) reported very positive responses to CRM training among a large sample of over 20,000 military and civilian aviation crew members, as well as significant overall increases in positive attitudes towards CRM skills after initial training (Helmreich & Wilhelm, 1991). Significant changes in behaviours targeted by the training were also reported when CRM training was combined with LOFT, i.e. simulation training in civil aviation (Helmreich & Foushee, 2010). Similarly, in military aviation CRM training has been shown to result in improved teamwork skills (Salas, Fowlkes, Stout, Milanovich, & Prince, 1999). Evidence for the impact of CRM training on flight safety, however, is largely lacking, which may in part be due to the extremely low incidence of accidents, and the continuous development of aviation technology over the last three decades resulting in technology-related decreases in accident rates that likely confounded longitudinal data (Helmreich & Foushee, 2010). Many anecdotal reports and retrospective analyses of severe incidents, however, indicated that concepts taught in CRM programs have helped to mitigate error and prevent catastrophic outcome (Helmreich & Foushee, 2010). The most recent and prominent example of effective leadership, decision making, communication and team coordination was the successful emergency landing of U.S. Airways Flight 1549 in the Hudson River in 2009 (Eisen & Savel, 2009).
While CRM training has been widely implemented and accepted as a successful strategy to improve safety and control error, several problems were identified during the first three decades of its use in aviation. A notable finding that has been consistently reported in studies evaluating CRM training effectiveness both in military and civilian aviation is the presence of a small percentage of trainees who reject the concepts of CRM (Helmreich & Wilhelm, 1991). These individuals, who have been referred to as boomerangs, cowboys, or drongos in aviation jargon, were identified by negative changes in attitudes towards CRM skills, i.e. these individuals had worse attitudes about CRM after a training intervention than they did before the training (Helmreich & Wilhelm, 1991). Individual factors such as personality and group dynamics have been associated with the boomerang effect. In a cluster analysis of personality inventories completed by pilots, Chidester et al. (1991) identified three common clusters. The first cluster represented pilots with high levels of positive instrumental traits and expressiveness, i.e., goal orientation in conjunction with high interpersonal orientation, and was referred to as “Positive Instrumental/Expressive”, or the “right stuff”; the second cluster comprised high levels of negative instrumentality, verbal aggressiveness and low interpersonal or team-orientation (e.g., competitiveness), as observed in “lone-wolf”-pilots with an autocratic leadership style, which was termed “Negative Instrumental”, or the “wrong stuff” (Chidester, Helmreich, Gregorich, & Geis, 1991). The third cluster of traits was characterized by the lack of either group of traits, instrumental and expressive, which was termed “Low Motivation”, or “no stuff” (Chidester et al., 1991). Chidester et al. (1991) studied the changes in pilots’ attitudes toward CRM-skills before and after CRM training and found pilots with the “right stuff” to benefit most from such training, while the “Low Motivation” or “no stuff” group had the least benefit or even showed a negative change in attitudes, as observed in the “boomerangs” (Chidester et al., 1991). Helmreich and Wilhelm (1991) considered “boomerang” pilots a safety threat and advocated for organizational interventions to remediate individuals who are resistant to CRM training. Moreover, the authors concluded that personality factors influenced both innate attitudes and the effect of CRM training and should, therefore, be considered in the development of training interventions, as well as the selection of suitable candidates for pilot training programs (Helmreich & Wilhelm, 1991).
A further important finding from the aviation literature is the fact that training effects, both at the level of attitudes and behaviours, decreased over time if no recurrent training was administered, necessitating recurrent CRM training in regular intervals (Helmreich et al., 1999). Lastly, Helmreich and Foushee (2010) underscored the importance of organizational support and leadership buy-in to ensure the success of CRM training programs (Helmreich & Foushee, 2010). Today, the training and assessment of nontechnical and CRM skills is an integral part of military and commercial aviation worldwide. Regulatory bodies such as the JAA, overseeing civil aviation in European member countries, the Civil Aviation Authority (CAA) in the United Kingdom (U.K.), and the FAA in the U.S., have mandated CRM-training for all operators and provide in their regulations clear instructions regarding initial CRM-training for new members of the flight crew, as well as instructions regarding recurrent training and assessment of nontechnical skills (JAA, 2007; CAA, 2013). For instance, in their operating requirements for domestic flights in the U.S., the FAA states that “After March 19, 1998, no certificate holder may use a person as a flight crewmember, and after March 19, 1999, no certificate holder may use a person as a flight attendant or aircraft dispatcher unless that person has completed approved crew resource management (CRM) or dispatcher resource management (DRM) initial training, as applicable, with that certificate holder or with another certificate holder” (Title 14, Chapter I, Part 121, subpart N, §121.404) (FAA, 1996).

Since its first definition and proposal during the NASA industry workshop in 1979 and the subsequent implementation in military and commercial aviation in the early 1980s, CRM-training has been adopted as a strategy to enhance safety by many other high risk, high reliability industries such as offshore oil production, nuclear power generation, and health care (Flin, O’Connor, & Mearns, 2002). The evolution of CRM and nontechnical skills in healthcare will be discussed in the subsequent chapters.

1.1.2 Adverse Events and Human Error in Operative Patient Care

Primum non nocere, this Latin phrase of unknown origin represents a fundamental principle in medicine: First, do no harm (Smith, 2005). Unintended injury to patients has been a major
concern in modern medicine. In 1961, an investigation at the Yale University Medical Service into complications affecting hospitalized patients reported a 20 percent incidence of iatrogenic injuries (Schimmel, 1964). Following the example of aviation, research efforts in healthcare into causes of these injuries began with studies of incidents, adverse events, and adverse patient outcome. The OR as a high-risk work environment within healthcare was the subject of the first studies, which were pioneered by anaesthesiologists in the U.S. as early as 1961. In a retrospective analysis of 33,224 cases of surgical patients that had received either spinal anaesthesia or general anaesthesia with muscle relaxants in a 10 year period at one hospital, Dripps et al. (1961) reviewed charts, incident reports and autopsy reports of all patients that had died within a 30 day postoperative period (Dripps, Lamont, & Eckenhoff, 1961). 80 deaths were deemed to be directly related to anaesthesia and scrutinized regarding factors causing or contributing to the adverse outcome (Dripps et al., 1961). Overall, 94 percent of the deaths were related to factors that with today’s knowledge would be classified as human factors, including inexperience (59 percent), inadequate preoperative preparation (19 percent), errors in judgment (6 percent), inadequate postoperative observation (7.5 percent), and inadequate diagnosis and management of haemorrhage (2.5 percent) (Dripps et al., 1961). Although Dripps et al. (1961) had conducted the study to investigate mortality rates associated with the then-new spinal anaesthesia and general anaesthesia with muscle relaxants and did not expressly discuss the topic “human factors”, they nevertheless commented on the observed human failures: “Preparation for anesthesia and operation is the responsibility of all concerned with the management of a patient. The anesthetist must share blame if preparation is inadequate. He is a physician, not a technician. He must not reject the responsibility by saying ‘that’s the surgeon’s job’” (p.110) (Dripps et al., 1961).

Subsequent studies targeting root-causes of adverse events differentiated between human and equipment failure. In a study investigating causes of critical incidents in anaesthesiology, Cooper at al. (1978) conducted forty-seven interviews with anaesthesiologists, asking about preventable incidents they had witnessed or experienced first-hand (J. B. Cooper, Newbower, Long, & McPeek, 1978). Using a standardized interviewing technique, interviewees were asked for details about the recalled incidents, including contributing factors. From these interviews, the authors identified and coded 359 incidents that were deemed preventable by the interviewees, ranging from equipment
malfunctions with no further consequence for the patient, to severe incidents that resulted in patient death (e.g., breathing-circuit disconnections) (J. B. Cooper et al., 1978).

Interestingly, human error was involved in 82 percent of the incidents, while equipment failure was identified as sole cause in only 14 percent (J. B. Cooper et al., 1978). The most frequently observed examples of human error were inadequate experience (77 percent), lack of familiarity with equipment (45 percent), poor communication with the team (27 percent), haste and inattention to detail (both 26 percent), fatigue (26 percent), and failure to perform checks (22 percent). In an Australian report of anaesthetic mishaps, Craig and Wilson (1981) analyzed 81 cases collected through anonymous incident reporting, asking for details and factors associated with the events. Of the 81 reported events, 65 percent were identified by the authors as being related to human factors, while a combination of human and equipment factors was found to have caused 12 percent of the mishaps (J. Craig & Wilson, 1981). The most commonly identified human failures causing incidents were the failure to perform normal checks (33 percent), distraction and inattention (12 percent each), hurry (11 percent) and fatigue (9 percent) (J. Craig & Wilson, 1981). From their results, the authors suggested implementing a “cockpit drill” in the beginning of routine cases to avoid or capture errors related to the failure to perform normal checks (J. Craig & Wilson, 1981). Building on findings and hypotheses from previous studies, Kumar et al. (1988) conducted a prospective survey of anaesthesiology staff members at one hospital regarding critical intraoperative incidents. The authors defined critical incidents as “an incident or mistake which could be harmful or potentially harmful to the patient during management of anaesthesia” (p.879) (Kumar, Barcellos, Mehta, & Carter, 1988). Incident report forms required participants to describe the event and the resulting outcome, provide a recommendation to prevent the event, and to indicate factors associated with the event on a checklist (Kumar et al., 1988). The items on the checklist were largely based on the findings that had been reported by Cooper et al. (1978), as described above. In addition to previous purely descriptive studies, Kumar et al. (1988) collected the incident report forms during two consecutive nine-month periods in 1984 and 1985, respectively. At the end of the first data collection period, the authors introduced an equipment checklist to be used prior to administration of each anaesthetic (Kumar et al., 1988). The results of the study confirmed previous reports in that during both observation periods 80 percent of critical incidents were directly linked to
human error, while only 20 percent were solely the result of equipment failure (Kumar et al., 1988). Similar to findings from previous studies, the most commonly reported causes of events were failures to perform normal checks, lack of experience or lack of familiarity with equipment, inattention or carelessness, distraction, and haste (Kumar et al., 1988). Interestingly, the number of reported critical incidents was halved after implementation of the equipment checklist, to 43 incidents in the second survey period, from 86 incidents in the first period, highlighting the effectiveness of checklists to prevent and capture errors (Kumar et al., 1988).

While the majority of studies had focused on a single hospital or preselected patient sample, the Harvard Medical Practice Study in 1991 was one of the first population-based, large-scale studies to investigate the rate of adverse events associated with treatment in acute care hospitals (Brennan et al., 1991). By means of retrospective chart review, the authors selected a random sample from all 2,672,863 patients that had been treated in acute care hospitals in the State of New York in 1984, focusing on adverse events associated with hospital care, as well as associated hospital, physician and patient factors (Brennan et al., 1991). The study group analyzed a sample of 30,195 patients that had been hospitalized in 1984, identifying 1,278 adverse events of which 306 were deemed to be due to negligence (Brennan et al., 1991). These numbers translated to a state wide adverse event rate of 3.7 percent, 28 percent of which were thought to be due to negligence (Brennan et al., 1991). In the second part of the Harvard study, Leape et al. (1991) analyzed the data set regarding adverse events by area of care the events occurred, and found that 41 percent of the adverse events occurred in the OR. Moreover, of all adverse events related to errors in operative care, 17 percent were attributed to negligence (Leape et al., 1991). In a more detailed analysis of the results, which was published in 1994 in a book chapter, Lucian Leape (1994b) differentiated between diagnostic errors (e.g., errors or delays in diagnosis, or the failure to respond appropriately to test results), treatment errors (e.g., errors in the technical execution of a procedure, avoidable delays in administering treatment, or unnecessary treatments), and preventive errors (e.g., omission of required prophylaxis, failure to monitor or follow up) (Leape, 1994b). A very important finding from the results of the Harvard study was the fact that 70 percent of all adverse events that affected patients were caused by human error and, therefore, potentially preventable (Leape, 1994b). When looking specifically at adverse
events related to operative care, 74 percent were considered to be preventable (Leape, 1994b).

Similar results were reported by Gawande et al. (1999), who screened a random sample of 15,000 admissions to Colorado and Utah hospitals from the year 1992 looking for adverse events in surgical care occurring within 30 days after an operation. Gawande et al. (1999) found that surgical adverse events accounted for 66 percent of all adverse events in hospitalized patients and had an annual incidence of 1.9 percent of all hospital admissions in 1992. When the location of the adverse event was taken into account, 40 percent were found to occur in the OR, and 17 percent of those were deemed the result of negligence (Thomas et al., 2000). Overall, the authors concluded that 54 percent of identified surgical adverse events could have been prevented (Gawande, Thomas, Zinner, & Brennan, 1999).

In the Canadian Adverse Event Study, Baker et al. (2004) screened 3,745 charts that were randomly selected from hospitals across Canada, calculating an overall adverse event rate of 7.5 percent. Similar to previous studies, 41.6 percent of adverse events were judged to be preventable, and the majority (34 percent) were linked to surgical procedures (G. R. Baker et al., 2004). The numbers from the Canadian study were corroborated by a systematic review of the incidence of adverse events in hospitalized patients (de Vries, Ramrattan, Smorenburg, Gouma, & Boermeester, 2008). Using combined data from eight studies comprising 74,485 patient records from hospitals across the U.S., Canada, the U.K., Australia and New Zealand, De Vries et al. (2008) calculated a median incidence of adverse events of 9 percent, with 40 percent being related to an operation. Similar to previous reports, 43.5 percent of adverse events were found to be preventable.

The majority of studies discussed in this chapter thus far examined incidences and root-causes of in-hospital adverse events. The evidence indicates that 1), a large proportion of adverse events are associated with operative care, allowing the conclusion that, within healthcare, the OR is a high-risk work environment; 2), over 40 percent of adverse events across studies are deemed to be caused by human error and, therefore, potentially preventable; and 3), failures in certain skills and behaviours (e.g., communication, inattention and distraction, and the failure to perform normal checks) have been directly
associated with adverse events, supporting the training of those skills to enhance safety.

Remarkably, despite event rates that were substantially higher than in aviation or other high-risk industries, error in medicine did not receive much public attention for a long time. One possible explanation for the lack of public attention, according to Leape (1994a), was the poor visibility of medical adverse events and deaths. For instance, if the data from the Harvard study were extrapolated to the U.S., the number of people dying of adverse events in operative care would reach 236 every day (Leape, 1994b). The same number of deaths in a single day would receive immediate worldwide coverage in the media if caused by a plane crash (Leape, 1994a). A further factor that has been impeding effective error management and the development of preventive measures in medicine for a long time is the professional culture of medicine, in which error is viewed as a personal failure, and a cause for shame (Leape, 1994a). This notion is reflected in the “person approach” to human error, which is widely practiced in healthcare and results in the identification and punishment of an individual immediately responsible for an adverse event. Fear of punishment and blame, together with the fear of litigation, leads to underreporting of critical incidents which, if analyzed appropriately, may have provided valuable information for the prevention of similar events in the future (Leape, 1994a). Based on James Reason’s (1990) conception of human error in complex systems, Leape (1994b) advocated the system approach to human error, creating a non-punitive environment to report and analyze error and near-miss data to create systems that avoid, trap or mitigate error, as practiced in aviation. The system approach to error was subsequently endorsed by Kohn et al. (2000) who, in their 1999 seminal report “To err is human” on patient safety in the American health care system, highlighted the unacceptably high rate of adverse events associated with human failure. In their report, the authors made recommendations to improve patient safety using proven approaches and strategies from aviation including briefing, checklists, and CRM training to enhance interdisciplinary teamwork and communication, and thus reduce error rates (Kohn LT, 2000).

After the publication of “To Err is Human” by the Institute of Medicine (Kohn LT, 2000), several research groups started looking into adaptive strategies to reduce or mitigate error in the OR, similar to CRM training in aviation. Behaviours, skills and attitudes that could potentially enhance or endanger patient safety were in the focus of the observations. In a
study to explore the influence of human factors on patient outcome. De Leval et al. (2000) conducted a prospective observational study of OR teams during 173 highly complex pediatric cardiac operations in 16 centres in the U.K. The complexity and high mortality (6.6 percent) of these operations allowed the authors to explore the relationship between human error and patient outcome. Observers gathered data regarding intraoperative events, classified as either minor (i.e., disruptions of the surgical flow without immediate risk to the safety of the patient), or major events (i.e., events posing immediate threat to the safety or life of the patient). Observers further judged whether the events were compensated by the team or remained uncompensated (de Leval, Carthey, Wright, Farewell, & Reason, 2000). In an analysis of the relationship between the observed variables (major and minor events; compensated or uncompensated) and the measures of patient outcome, de Leval et al. (2000) found the number of major events per case to be a strong predictor of the outcome variables “death” and “death and/or near miss”. Minor events, although not directly causing adverse patient outcome, were found to have a multiplicative effect, with a significant positive correlation of the number of minor events and patient death and/or near miss (de Leval et al., 2000). When accounting for compensation of events by the surgical teams the authors found a significant reduction in the risk of death for compensated major events when compared with uncompensated major events (de Leval et al., 2000). Observed examples of surgeons’ behaviours associated with the compensation of events were described by the authors as diagnostic skills, knowledge of surgical problem-solving strategies, and communication with the team (de Leval et al., 2000).

Carthey et al. (2003) further analyzed the data generated by De Leval and colleagues (2000) to explore the relationship between surgeons’ nontechnical skills and patient outcome. Human factors data were gathered during the observations using a framework of behavioural markers of surgical excellence that the group had developed based on aviation data, grouping behaviours into individual level markers (e.g., mental readiness, cognitive flexibility, anticipation, safety- and situation awareness, communication style); team level markers (e.g., team leadership, communication between team members, coordination between team members); and organizational level markers (e.g., policy; planning and scheduling; organizational culture) (Carthey, de Leval, Wright, Farewell, & Reason, 2003). Surgeons were grouped according to a score of procedural excellence, calculated from the
number of minor and major events, as well as patient death and death and/or near miss, showing that surgeons with the best scores of procedural excellence exhibited more of the behavioural markers than surgeons with lower scores (Carthey et al., 2003). The results supported the role of nontechnical skills in the avoidance, trapping and mitigation of errors and events, with the potential of preventing adverse patient outcome.

Catchpole et al. (2007) applied a similar methodology to identify and analyze failures in successful operations. A human factors expert observed 24 pediatric cardiac and 18 orthopaedic operations, recording intraoperative failures and events and rating nontechnical performance of the OR teams (K. R. Catchpole et al., 2007). Intraoperative events were classified according to three degrees of severity, with minor problems representing undesired events with no immediate impact on the operation (e.g., distractions), operating problems that disrupted the operative flow, and major problems that potentially endangered patient safety (K. R. Catchpole et al., 2007). Nontechnical performance was assessed using a modified version of the aviation NOTECHS, and rated in four categories, based on observable behaviour: leadership and management; teamwork and co-operation; problem solving and decision making; and situation awareness (K. R. Catchpole et al., 2007). Based on the proportion of high and low scores, OR teams were classified in two groups: effective teams, and ineffective teams (K. R. Catchpole et al., 2007). Minor problems occurred at an average rate of 9.5 (pediatrics) and 13.1 (orthopaedics) per case and were inversely correlated with operative time, with each minor problem causing a delay of, on average, 3 minutes and 36 seconds (K. R. Catchpole et al., 2007). Operating problems that disrupted flow and resulted in increased operative time occurred at a rate of 5.7 per case in paediatric, and 5 per case in orthopaedic surgery (K. R. Catchpole et al., 2007). In operations with at least one major intraoperative event, Catchpole et al. (2007) identified significantly more minor failures when compared to operations without major failures, corroborating the findings of de Leval and colleagues (2000) that even seemingly harmless events can accumulate to endanger patient safety. Also similar to de Leval et al.’s (2000) results, problems with coordination and communication were judged to be among the most common causes for minor events, and were the single most common category of factors associated with flow disruptions (K. R. Catchpole et al., 2007). The analysis of NOTECHS scores of OR teams confirmed these findings, with the group of effective teams outperforming the
ineffective teams with regards to intraoperative performance, and operative time (K. R. Catchpole et al., 2007). Based on their observations, Catchpole et al. (2007) recommended nontechnical skills training for OR teams to avoid, trap and mitigate minor events and potentially prevent more serious events from causing harm to the patient (K. R. Catchpole et al., 2007).

1.2 Nontechnical Skills in Surgery

In the previous chapter, I have summarized the history of nontechnical skills in aviation, which were first described as specific attitudes, skills and behaviours of flight crews associated with safe flight operations, and were subsequently addressed in targeted training interventions termed CRM training. I have outlined the evolution of CRM training as a strategy to improve team performance, avoid, trap and mitigate error, and enhance safety, which subsequently led to its worldwide implementation in commercial and military aviation, as well as other high-risk industries. In a review of adverse events in healthcare in general and the OR in particular I then presented evidence supporting the critical role of human error and nontechnical skills in these, often preventable, events. Next, I will provide a more detailed account of nontechnical skills relevant to the OR environment, discuss the pertinent evidence supporting these skills and behaviours, and review previous interventions targeted at the improvement of nontechnical performance in the context of the OR.

1.2.1 Terminology and Definition

The term “nontechnical skills” encompasses a variety of knowledge, skills, behaviours and attitudes that, in the context of surgeons in the OR, are the subject of this thesis. As a generic term, “nontechnical skills” does not describe very well the different aspects of performance it entails. This is reflected in the definition of the adjective “non-technical” provided in the Oxford English Dictionary (OED) (“non-technical”, 2014) as “Not relating to or involving science or technology; not requiring or assuming specialized or technical knowledge”, which may imply that skills described by this adjective are less relevant and do not need to
be learned or practiced (Nestel, Walker, Simon, Aggarwal, & Andreatta, 2011). Quotations provided by the OED to exemplify the use of the word completely lack examples in the context used for this thesis (“non-technical”, 2014). This is rather unfortunate, as the adjective nontechnical has been used for almost 20 years to describe a skill set that had previously been subsumed under the term “CRM skills” (van Avermaete, 1998).

Nontechanical skills are occasionally referred to as human factors in the literature, which I believe is not entirely accurate. Although nontechnical skills are a subset of human factors, the larger construct of human factors entails a wide variety of interactions in a work environment, including those at the human-machine or human-technology interface such as design aspects (e.g., usability and handling, safety) of equipment, and other issues pertaining to the domains of engineering and industrial design, that are also referred to as ergonomics (Helmreich & Davies, 1996). In an aviation context, nontechnical skills have been defined as “the cognitive and social skills of flight crew members in the cockpit, not directly related to aircraft control, system management, and standard operating procedures” (Flin et al., 2003). Providing a more generic, context-independent definition, Professor Rhona Flin of the University of Aberdeen previously described nontechnical skills as “the cognitive, social and personal resource skills that complement technical skills, and contribute to safe and efficient task performance” (p.1) (Flin, O'Connor, & Crichton, 2008). More recently, Gordon et al. (2014) conducted a Delphi study to achieve consensus on the definition of nontechnical skills in healthcare, proposing the following definition: “A set of social (communication and team working) and cognitive (analytical and personal behaviour) skills that support high quality, safe, effective and efficient multiprofessional care within the complex healthcare system”. For the purposes of this thesis, I will use the definitions of nontechnical skills by Flin et al. (2008) and Gordon et al. (2014), and I will use the term “nontechnical skills” synonymous with “CRM skills”, which is still used by some researchers to describe the same set of attitudes and skills as well as behaviours.

1.2.1.1 Definition and Characterization of Nontechnical Skills

In this paragraph, I will introduce a selection of nontechnical skills that are relevant for surgeons in the OR and will thus be the subject of the present work. Since a universally
accepted taxonomy or catalogue of nontechnical skills in the OR is lacking, I have based the selection of attitudes, skills and behaviours for this thesis on published frameworks and evidence in the surgical literature. Thus, the resulting selection should not be regarded as an exhaustive list of nontechnical skills in the OR. Rather, it reflects the current state of knowledge and opinions of scholars in this area of research. A detailed description of existing frameworks and taxonomies of nontechnical skills will be provided in the paragraph on assessment later in this chapter. To enhance clarity, I will present nontechnical skills grouped, based on previous empirical research in aviation and in the OR environment, as well as basic conceptions of social and cognitive psychology. Howard et al. (1992), for instance, defined key principles of CRM in the OR based on aviation CRM skills, coining the term Crisis Resource Management. In their definition of CRM, the authors emphasized two main categories of skills: Decision making and cognition; and teamwork and resource management (Howard, Gaba, Fish, Yang, & Sarnquist, 1992). The former encompassed skills and behaviours such as knowing the environment, anticipating and planning, using available information, cross-checking, and preventing fixation errors (Howard et al., 1992). The latter included behaviours such as exercising leadership, calling for help, communicating effectively, distributing workload, and using all available resources (Howard et al., 1992). Based on extensive work in aviation, other high-risk industries and healthcare, Flin et al. (2008) distinguished between cognitive skills (i.e., situation awareness and decision making), social skills (i.e., communication, teamwork, and leadership), and personal resource skills (managing stress, and coping with fatigue). Gordon et al. (2014), in their recent Delphi survey to achieve consensus on nontechnical skills for education in healthcare, categorized skills as either social or cognitive. In their competency framework, however, the authors differentiated between cognitive skills pertaining to personal behaviours and those representing analytical skills (Gordon, Baker, Catchpole, Darbyshire, & Schocken, 2014). Personal behaviours, according to Gordon et al. (2014), comprise attributes such as compassion, integrity and honesty, but also behaviours pertaining to the identification and management of stress and fatigue, while analytical skills refer to situation awareness and decision making.

To enhance clarity, I will subdivide nontechnical skills into three main conceptual groups: 1) cognitive skills, comprising cognitive functioning and processing of thoughts within an
individual; 2) social or interpersonal skills that are used to interact with other individuals when working in a team; and 3) personal behaviours and personal resource skills that form a heterogeneous group of knowledge, skills and behaviours, and attributes that have been associated with safety in high-risk environments, or are deemed important attributes and skill of surgeons in the OR.

1.2.1.1 Cognitive skills

In the *Glossary of Psychological Terms* on the webpage of the American Psychological Association (APA) cognitive processes are defined as “*higher mental processes, such as perception, memory, language, problem solving, and abstract thinking*” (“Cognitive processes,” n.d.). Of the cognitive processes in the APA definition, the skills pertaining to “perception” and “problem solving” have gained the greatest significance in the field of human factors research, both in aviation and health care. These skills, or rather skill categories, are situation awareness and decision making. Both situation awareness and decision making, the latter of which entails processes related to problem solving, are closely related but clearly represent separate constructs and will, therefore, be discussed separately.

Situation awareness can be described as “knowing what’s going on”, or, adapted from Mica Endsley (1988), a surgeon’s internal model of the environment around him or her at any point in time. A more accurate and likely the most widely accepted definition of situation awareness has been proposed by Endsley (1995): “Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future” (p.36) (Figure 2).
At the first level, *perception*, information is perceived from the environment, both passively by detecting cues and actively by gathering information (Endsley, 1995). At the second level, *comprehension*, the cues perceived and gathered from the environment are compared with existing information and assessed in view of current goals and processes to understand
their meaning for the situation at hand (Endsley, 1995). At the third and highest level of situation awareness, *projection*, information is analyzed and interpreted regarding potential consequences for the system in the near future (Endsley, 1995). It is important to know that this model, although not context specific, conceptualizes situation awareness in a specific environment at one particular point in time (Endsley, 1995). This temporal aspect of situation awareness, as well as the need for constant assessments to maintain good situation awareness, has also been highlighted by other authors (Sarter & Woods, 1991). The third level for instance, projecting the future state, applies to the immediate future of the current environment. In military aviation, as a highly complex environment with rapidly changing conditions, pilots must obtain and maintain good situation awareness at all times to ensure a safe flight, and successfully complete their mission (Endsley, 1995). In a combat situation, for instance, a pilot must monitor parameters of the own aircraft (e.g., altitude, airspeed, tank content), enemy aircraft (e.g., speed, position, vector), and ambient conditions such as weather, among many other factors. The pilot must then analyze the relevance of all factors in view of the current situation to understand what is going on, and to project the status of the environment (e.g., position of own plane) in the near future (Endsley, 1995). These assessments have to be repeated constantly, as aircraft move very fast and conditions change rapidly. Consequently, pilots’ situation awareness has been recognized by some as the most important determinant of successful missions in military aviation (Endsley, 1988). Situation awareness, according to Endsley (1995), is not a static system; rather, it is influenced by external or system factors, as well as internal or individual factors. Examples of system factors are workload and stress, as both high workload and stress can affect the perception and comprehension of information (Endsley, 1995). Individual factors comprise current objectives and expectations, also referred to as mental model, which at the first level may direct the attention to specific cues that are relevant for the situation, and at the second and third level are used to interpret findings with regards to their relevance for the current situation (Endsley, 1995). The interpretation of findings at the projection level is also informed by an individual’s experience, for instance through recognition of familiar situations or combinations of factors that in previous encounters have led to adverse events (Endsley, 1995). It becomes clear that situation awareness is intimately connected with decision making, as the quality of a decision in any situation depends on the correct
assessment and understanding of the environment, as well as the projection of the future state in mental simulations. Moreover, as shown in Figure 2, any decision that is made in a particular environment has a direct impact on that environment, requiring constant reassessment in order to maintain good situation awareness. The OR, although not directly comparable to a fighter jet, is equally a dynamic, high risk environment where conditions can change quickly and often unexpectedly, requiring good situation awareness of all team members at any point in time. Moreover, many internal and external factors in the OR hinder the attainment or maintenance of good situation awareness. Examples for internal factors are stress and fatigue, and fixation on the operative field (“tunnel vision”). External factors include noise from conversations, music, and monitors, as well as physical barriers (Helmreich & Schaefer, 1994). Helmreich and Schaefer (1994) for instance described the sterile drape between the surgical team and the anaesthesiologist as a significant barrier to situation awareness of both teams, as it prevents both sides from perceiving important information about the patient. The relevance of situation awareness for patient safety in the OR has been highlighted in several previous studies. Errors in perception have been found to contribute to critical incidents in anaesthesia, as discussed in a previous paragraph of this chapter (J. B. Cooper et al., 1978). Similarly, errors in the perception and comprehension of problems were associated with surgical mortality as a result of anaesthesia (Dripps et al., 1961). Gaba et al. (1995) highlighted the role of situation awareness in anaesthesiology in a review of their own and others’ experiences from observations in the OR and in simulations. The authors identified several factors that impeded the perception and comprehension of information in the OR including: subtlety of cues; atypical patterns of cues; or multiple simultaneous cues resulting in information overload (Gaba, Howard, & Small, 1995). Gaba et al. (1995) also recognized the role of experience in the recognition and interpretation of cues and problems, but emphasized the need of constant re-evaluations to avoid fixation errors (e.g., continuing with a plan of action despite a lack of success). Although the above-mentioned studies focused on anaesthesiologists, I believe that the concept of situation awareness can be extrapolated to surgeons in the OR.

Recent studies support the relevance of surgeons’ situation awareness for intraoperative patient safety. Mishra et al. (2008) observed 26 laparoscopic cholecystectomies and assessed technical and nontechnical performance of the surgical team. The authors found situation
awareness of the surgeons, as measured with a modification of the aviation NOTECHS behaviour marker system (surgical NOTECHS), to be inversely correlated with the rate of technical errors during the operations suggesting that good situation awareness may help to prevent or trap errors (Mishra, Catchpole, Dale, & McCulloch, 2008). Poor situation awareness has also been shown to contribute to errors in intraoperative decision making. For instance, Way et al. (2003) analyzed 252 bile duct injuries resulting from laparoscopic cholecystectomies and identified errors in perception, i.e., situation awareness, to be the primary cause in 97 percent of the mistakes. Through a review of the operative reports and video recordings that were available for some of the operations the authors found that the injury was not detected intraoperatively in 75 percent of cases, although cues had been present in many of these situations (Way et al., 2003). Although the study by Way et al. (2003) was limited to the analysis of a single, very specialized operative procedure in a defined anatomic region, I do believe that the results underscore the importance of surgeons’ situation awareness in the OR. One important aspect of situation awareness that has not yet been addressed in the present paragraph is the concept of team situation awareness, which differs significantly from the model of individual situation awareness. While situation awareness of the individual is predominantly a cognitive process, team situation awareness involves both individual situation awareness of each individual team member, as well as team processes such as coordination, communication, and information sharing (Endsley, 1995; Prince & Salas, 1993; Salas et al., 1995). As such, team situation awareness represents a dynamic process requiring constant reassessment and updates to ensure good situation awareness of the entire team at any given point in time (Salas et al., 1995). Based on Endsley’s (1995) theory, the ability to acquire and maintain situation awareness may be influenced by training. Behaviours and skills that may be amenable to educational interventions include vigilance, active gathering of information in view of a problem, but also the constant reassessment of the situation to identify irregularities and trap errors. Gaba et al. (1995) proposed the training of several aspects of situation awareness, including scanning the environment to detect even subtle cues; allocating attention to the most pertinent problem in situations with multiple simultaneous cues; and recognizing patterns of cues. With regards to team situation awareness, Salas et al. (1995) recommended the training of both individual situation awareness through the practice of information-seeking
and processing behaviours, as well as behaviours to improve the sharing of information, such as communication and team planning. The interpersonal skills associated with teamwork, however, will be discussed in the paragraph on social skills.

The following paragraph deals with the second cognitive skill to be addressed here: decision making. Decision making has been studied extensively in aviation, where pilots’ decisions in critical situations can mean the difference between success and failure of a mission. Similarly, research efforts in anaesthesiology have targeted intraoperative decision making from the perspective of the anaesthesiologist, with a special focus on the identification and prevention of errors (Gaba, 1989). In surgery, decision making has only recently moved in the focus of attention, and especially the nontechnical or cognitive aspect of it is not very well understood (Flin, Youngson, & Yule, 2007). Therefore, I will begin this section with a review of the aviation and anaesthesiology literature before moving to surgeons’ intraoperative decision making.

Since the early 1990s, decision research in aviation, as well as other high-risk work environments, has focused predominantly on naturalistic decision making which, in contrast to traditional decision making with the goal of identifying an optimal solution to a problem, describes decision making in real-world settings involving high risk and time pressure (Orasanu, 2010). Klein (1997) defined naturalistic decision making as “the study of how people use their experience to make decisions in field settings” (p. 340). Naturalistic decision making thus concerns situations that often do not allow the meticulous analysis and comparison of all available options to identify the ideal solution to a problem, as done in traditional analytical decision making (G. Klein, 1997). As such, the concept of naturalistic decision making lends itself well to intraoperative decisions made by anaesthesiologists and surgeons. In the following, I will describe four modes of decision making that have been associated with naturalistic decision research: recognition-primed decision making; rule-based decision making; analytical decision making; and creative decision making.

Recognition-primed or intuitive decision making has been conceptualized by Gary Klein (1998) based on his field observations of firefighters, and comprises the assessment of the situation by the decision maker, the recognition of the problem from experience (e.g., through recognition of typical patterns of cues), the rapid evaluation of different courses of
action by imagining the outcome in mental simulations, and the implementation of the most appropriate course of action. This model of decision making is characteristic for experts in a field, as both the recognition of the problem from patterns of cues, and the mental simulation of the outcomes of different courses of action require a substantial amount of knowledge and previous experience to be recalled from long term memory (Gary Klein, 1998). Since recognition-primed decision making is intuitive and occurs almost automatic, it requires less cognitive resources (i.e., working memory) than the traditional analytical approach. A further strength of recognition-primed decision making is that it leads to a fast solution of a problem, which is often a requirement in high risk settings. Analytical decision making on the other hand involves the identification of a problem, the generation of options, comparison of options with regards to expected benefits and potential risks, and selection of the most appropriate option. It is almost a truism that the analytical approach requires substantial cognitive resources as well as time. Hammond et al. (1987) compared the efficacy of intuitive and analytical decision making in an observational study involving 21 expert highway-engineers in different tasks. The authors found that the intuitive approach was as effective as the analytical approach, and pointed out that it can even be superior in certain cases (Hammond, Hamm, Grassia, & Pearson, 1987). The authors further established, that the efficacy of either decision making approach depended on the nature of the task (Hammond et al., 1987). From their results, the authors proposed the cognitive continuum theory, describing decision making to occur on a continuum between intuition and analysis, depending on the nature of the task, which itself varies on a continuum between intuition-inducing (e.g., many cues, simultaneously displayed in short period of time) and analysis-inducing (e.g., few clues, sequentially displayed, long time) (Hammond et al., 1987). Hammond et al.’s (1987) results suggest that both intuitive and analytical strategies have their place in naturalistic decision making, depending on the context and the nature of the problem.

A different type of decision making, rule-based decision making, is characterized by the application of a stored rule, either from memory or from other sources (e.g., user’s manual, data base, textbook), to solve a problem (Rasmussen, 1983). Jens Rasmussen (1983), in his theory of cognitive control, differentiated between skill-based, rule-based and knowledge-
Based on the abovementioned theories of decision making, as well as aviation incident and accident reports, Orasanu and Fischer (1997) developed a model of decision making in aviation comprising of the two steps *situation assessment* and *selecting a course of action* (Figure 3). In their model, which is based on naturalistic decision theory, the decision making process begins with an assessment of the situation to identify the problem, and to judge urgency and potential consequences (Orasanu, 2010). The process of situation assessment in Orasanu’s (2010) model is very similar to Endsley’s (1995) model of situation awareness (Figure 2), again showing the close connection of situation awareness and decision making. In addition to the formal assessment of the situation, the problem can also be recognized by a specific pattern of cues, as described in recognition-primed decision making. The further course of the decision process then depends on the available time and the associated risk, as well as the familiarity of the problem (Figure 3). If the problem is not familiar or understood after the initial assessment, two options are possible: if enough time is available, the situation is reassessed; if time is scarce and the risk is high, the best fitting
rule is applied following the principle of satisficing as opposed to maximizing, which is characteristic of naturalistic decision making (Orasanu, 2010). Consequently, in high-risk high time pressure situations in aviation, a rule-based approach is most often selected (usually available in a flight manual) (Figure 3).

**Figure 3: Aviation decision process model.**


In situations where more time is available, decision makers can pursue an analytical approach by comparing several options and selecting the most appropriate one to solve the problem. The creative approach to decision making is available for familiar or known problems where an option is lacking (Figure 3). Flin et al. (2007) hypothesized that the
Aviation decision model could be modified to describe surgeons’ intraoperative decision making. In their model, the decision making process equally starts with an assessment of the situation. Depending on the risk and the time available to solve the problem, a course of action is chosen using either a recognition-primed or rule-based strategy for high-risk problems with limited availability of time, or an analytical or creative approach for problems with variable risk and more time (Flin, Youngson, et al., 2007). The validity of Flin et al.’s (2007) decision model, however, could not yet be confirmed in the context of surgical decisions in the OR.

Pauley et al. (2011) for instance conducted semi-structured interviews with 24 staff surgeons from three teaching hospitals in Scotland. Participating surgeons were asked to recall a critical intraoperative event that required a decision to be made. Responses were transcribed and, based on Flin et al.’s (2007) model, coded regarding the decision making approach used by the surgeon. Moreover, surgeons were asked to describe perceived threats and risk. Of all decision processes recalled by participants, the vast majority were either analytical (n=12, 50 percent) or recognition-primed (n=11, 46 percent). Only one surgeon recalled using a rule-based strategy (Pauley, Flin, Yule, & Youngson, 2011). Interestingly, types of operations, context (elective or emergency case), time pressure or threats were not associated with the choice of strategy. Rather, personal risk tolerance was found to predict the strategy used by surgeons, with statements about risk tolerance being more likely in surgeons selecting an analytic approach when compared with those making intuitive decisions (Pauley et al., 2011). These findings could be interpreted such that surgeons’ personality influences the choice of decision strategy with surgeons less willing to take risk favouring an analytic approach (Pauley et al., 2011). Pauley et al. (2013) subsequently applied their critical decision method to interview 12 ophtalmic surgeons in Scotland corroborating their previous findings. Half of the surgeons reported using an analytic approach when faced with an intraoperative problem, while the other half recalled using a recognition-primed method (Pauley, Flin, & Azuara-Blanco, 2013). Similar to the first study, Pauley et al. (2013) found risk tolerance to play an important role in decision making, although the authors did not quantify these findings.

In a different study, Jacklin et al. (2008) conducted semi-structured interviews with 10
general surgeons in the U.K., asking them to think about (and verbalize) the entire in-hospital treatment course of a patient with gallstones, including admission, surgery, postoperative care, and discharge. A special focus was directed on decisions at each step of the treatment course. After transcribing and coding the interviews, the authors identified two modes of decision making: an intuitive mode that appeared to be based on clinical experience; and a rule-based mode, where the interviewees had a personal rule for the particular situation (Jacklin, Sevdalis, Darzi, & Vincent, 2008). Analytical decision making was not mentioned, although the description of many situations contained at least elements of analytic decision making. For instance, the authors described a surgeon’s consideration of differential diagnoses as an example of intuitive decision making, although this cognitive process may also be interpreted as an example of analytical decision making. (Jacklin et al., 2008). The lack of evidence supporting Flin et al.’s (2007) model of decision making could be owing to the differences between surgery and aviation. For instance, the human body is too complex and variable to be described in a user’s manual as commonly used in airplanes.

The question as to how surgeons choose between different methods of decision making remains open. Pauley et al.’s (2011 and 2013) findings show an association of decision making with personal risk tolerance suggesting a role of surgeon personality. Other cognitive processes might also play a role. Moulton et al. (2010) recently described a phenomenon in surgeons they referred to as “slowing down” during phases of intraoperative difficulty. In semi-structured interviews with 28 surgeons, the authors identified four different stages of transition from automaticity to a more effortful state: fine-tuning; focusing more intensely; removing distractions; and stopping (Moulton, Regehr, Lingard, Merritt, & MacRae, 2010). The phenomenon of slowing down in an operation, or knowing when a more effortful mode should be assumed to cope with intraoperative difficulties, represents an important cognitive skill that likely plays an important role in intraoperative error prevention and patient safety. The slowing down moments may also influence decision making, for instance when changing between the rapid and automatic mode of recognition-primed decisions, to the more purposeful mode of analytic decision making. This relationship, however, has not yet been investigated.
In the context of human error and nontechnical skills a further important question is which decision making approach is the safest, or what can be done to reduce decision errors. Errors in decision making, which are referred to as mistakes (as opposed to slips or lapses) in James Reason’s (1990) error taxonomy, have been the topic of multiple studies in anaesthesiology. Examples of cognitive errors in decision making that have been identified in the anaesthesiology literature include confirmation bias (taking into account only cues and information that fit the diagnosis), and premature closure (quickly making a diagnosis without considering alternative options) (Vannucci & Kras, 2013). In aviation, an analysis of decision making errors revealed that 75 percent were due to plan continuation errors, i.e., adhering to a plan although several cues suggest it may be wrong (Orasanu, 2010).

Misinterpretation of cues or situations can also play a role in wrong decisions, as Way et al. (2003) have shown in their study about bile duct injuries, where the assumed recognition of anatomic cues led experts to make the wrong recognition-primed decision of transecting the common bile duct. Flin et al. (2007) suggested that rule-based decision making may be preferentially used by less experienced surgeons, who have less cue patterns and automatic responses stored in long term memory. Many intraoperative problems, however, are not represented by treatment guidelines and intraoperative crises are often complex, requiring at least some degree of analytical decision making.

A strategy from aviation CRM training is team decision making, where the cognitive load of decisions is shared by involving team members in the process (Orasanu, 2010). Involving the team in decision making has the potential advantage of additional input from others’ experiences, and enhances the likelihood of trapping cognitive errors such as misperceptions. Team processes can also be utilized to trap decision errors through active monitoring of others’ performance and by challenging decisions that are perceived as wrong (Salas, Sims, & Burke, 2005). The team aspect was equally emphasized in the Non-Technical Skills for Surgeons (NOTSS) framework and taxonomy of surgeons’ nontechnical skills in the OR (Yule, Flin, Maran, Rowley, et al., 2008). Yule et al. (2008) described positive behaviours pertaining to decision making as discussing options with other team members, asking for input from team members to generate options, sharing the selected option with the team, explaining why an option has been selected, and updating the team on
progress after a decision has been implemented (Yule, Flin, Maran, Rowley, et al., 2008). Interpersonal skills of surgeons in the OR will be addressed in the following paragraph on social skills.

1.2.1.1.2 Social skills

Social or interpersonal skills describe skills needed to interact with other human beings. The importance of social skills in high-risk environments such as the cockpit has been identified early in root-cause analyses of adverse events, as elaborated in the first paragraph on nontechnical skills in aviation. Subject of the following paragraph are three social skills that have been identified as relevant for safety and efficiency in high-risk settings, either through human factors research in aviation, or observations directly in the OR: teamwork, communication, and leadership.

“Teamwork” is a familiar term in general linguistic usage, and can be described as the “cooperative effort by the members of a group or team to achieve a common goal” (“teamwork”, n.d.). Teamwork is ubiquitous in modern societies, and examples of teams can be found in sports, science, the corporate world, and healthcare. Salas et al. (1992) previously defined a team as “a distinguishable set of two or more people who interact dynamically, interdependently, and adaptively toward a common and valued goal/object/mission, who have been assigned specific roles or functions to perform, and who have a limited life span of membership” (p.4) (Salas, Dickinson, Converse, & Tannenbaum, 1992). In contrast to a group of individuals working together, a team has interdependent members each of who contribute a highly specialized skill set allowing them to share the workload. Moreover, a team is able to adapt to changing conditions in dynamic work environments. Salas and coworkers (2005) defined teamwork as a “set of interrelated thoughts, actions, and feelings of each team member that are needed to function as a team and that combine to facilitate coordinated, adaptive performance and task objectives resulting in value-added outcomes” (p. 562). In a review of teamwork literature, Salas et al. (2005) identified behaviours and attitudes that have been associated with effective team functioning, defining the “big five” components of teamwork: team leadership; mutual
performance monitoring; backup behaviour; adaptability; and team orientation (Salas et al., 2005). Briefly, team leadership comprises the coordination, planning and organization of team activities. Mutual performance monitoring describes an aspect of team situation awareness, and requires team members to have a common understanding of the overall plan in order to detect errors or deviations in performance. Backup behaviour describes team members’ ability to anticipate others’ needs and provide help accordingly. Adaptability refers to a team’s ability to effectively respond and adapt to changing conditions in the environment. Team orientation represents the notion of individuals in a team to value the team’s goals higher than their own goals for the duration of their team membership (Salas et al., 2005).

In addition to the five components of teamwork, Salas et al. (2005) identified “coordinating mechanisms” that had been associated with enhanced team performance and effectiveness. One of the mechanisms that help to coordinate teamwork is the concept of shared mental models (Salas et al., 2005). A shared mental model can be described as a partial overlap between team members’ understandings of each other’s duties, resources and capabilities pertaining to the case at hand (Salas et al., 2005). Clarity about roles in a team setting is also thought to be part of the shared mental model (Stout, Cannon-Bowers, Salas, & Milanovich, 1999). In the OR, for instance, surgeons, nursing staff, and anaesthesiologists represent highly specialized groups or sub-teams that need to work together as an OR team for the duration of an operation. In order to perform teamwork behaviours such as monitoring others’ performance and providing help and support, team members need a basic understanding about their teammates’ tasks, capabilities and resource requirements. This understanding can be attained by communication, e.g., asking teammates what their needs are and whether they need help. In situations of high workload and high stress, however, when communication can deteriorate, shared mental models provide an implicit understanding of the team processes, task requirements and goals that allows team members to coordinate team activities, monitor others’ performance and allocate resources to support them (Cannon-Bowers, Salas, & Converse, 1993). A formal method to establish or update a shared mental model is briefing, which will be discussed in more detail in the following paragraphs (Allard, Bleakley, Hobbs, & Vinnell, 2007). Evidence for the effectiveness of
shared mental models in improving team performance can be found in the psychology literature (Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000).

A further coordinating mechanism of teamwork according to Salas et al. (2005) is closed-loop communication. Closed-loop communication describes the process of an effective transmission of messages involving a sender and a receiver. The sender transmits a message, while the receiver confirms receipt and understanding of the message back to the sender (Salas et al., 2005). If receipt is not confirmed, it is the responsibility of the sender to check back with the receiver and request confirmation (Salas et al., 2005). The result is an effective exchange of information that ensures important messages are heard and understood, while preventing ambiguities and misunderstandings. The concept of closed-loop communication is not new. In aviation as well as the military, this form of communication has been a longstanding tradition to ensure the successful transmission of critical information (e.g., read-back of orders). Communication though will be discussed as a discrete social skill later in this paragraph.

For the purposes of the present work, as mentioned earlier, I conceptualize surgeons’ nontechnical skills after Flin et al. (2008) as a set of cognitive and social skills that enhance safety and efficiency in the OR. Therefore, for the purposes of the present thesis, I will distinguish between “teamwork” as an overarching concept as defined by Salas et al. (2005), and “teamwork skills” or “teamworking” as a category of social skills and behaviours that are enacted by surgeons to improve the cooperation with other team members in the OR setting. Consequently, leadership and communication, although essential for effective teamwork, will be discussed as separate nontechnical skills.

In the aviation NOTECHS framework of nontechnical skills, Flin et al. (2003) defined “cooperation” as an independent category of skill comprising behaviours such as team building, support of others, understanding team needs, and conflict solving. Similarly, Mishra et al. (2009) used the same skills for the category “teamwork and co-operation” in their framework of nontechnical skills of OR teams (Oxford NOTECHS). Fletcher et al. (2003), for their taxonomy of anaesthesiologists’ nontechnical skills (ANTS), included in
the skill category “teamworking” such skills as co-ordinating activities, exchanging information, and supporting others (Fletcher et al., 2003). Last, Yule et al. (2006), in their framework and taxonomy of surgeons’ nontechnical skills in the OR (NOTSS), combine teamwork and communication in a skill category described by the skill elements “exchanging information”, “establishing a shared understanding”, and “co-ordinating team activities”. For the purposes of the present work teamwork as a category of social skills will be defined as the support of others, co-operation with team members, co-ordination and synchronization of team tasks, and the creation and maintenance of a shared mental model (e.g. through team briefings and debriefings). The relevance of teamwork-related behaviours was previously demonstrated by Mazzocco and colleagues (2009), who conducted a prospective observational study to investigate the influence of intraoperative teamwork on postoperative patient outcome. In structured observations of 300 surgical procedures at 4 hospital sites, trained observers rated teamwork based on four behavioural marker categories that had previously been associated with effective teamwork: briefing, information sharing, inquiry, and vigilance and awareness (Mazzocco et al., 2009). The authors conducted multiple regression analyses to evaluate the relationship of the teamwork-related behaviours with 30-day postoperative patient outcome while accounting for patient factors such as comorbidities and individual risk factors (Mazzocco et al., 2009). The results showed that patients were at an increased risk to suffer minor or major complications or death, when certain team behaviours were less frequently observed (Mazzocco et al., 2009). Specifically, the odds for poor outcome were increased when information sharing behaviours were less frequently observed during the intraoperative phase and during handoffs, and when briefings were less frequently observed during handoffs (Mazzocco et al., 2009).

Communication is defined in the American Heritage Dictionary® of the English Language (“communication”, n.d.) as “The exchange of thoughts, messages, or information, as by speech, signals, writing, or behavior”. Communication is an essential interpersonal skill and its role in the causation and prevention of errors has been studied extensively in aviation and, more recently, in healthcare (Helmreich & Schaefer, 1994). Some scholars view communication not as an independent skill, but as a means to perform other social skills.
The NOTECHS framework of nontechnical skills in aviation, for instance, lacks a discrete category of communication (van Avermaete, 1998). Since this thesis is focused on problems pertaining to the teaching and assessment of nontechnical skills, I opted to maintain communication as a discrete skill as it comprises skills and behaviours that are potentially amenable to training and assessment. Communication in the OR has been among the first competencies to be targeted by human factors research in healthcare. From their own observations in the OR of a Swiss teaching hospital, Helmreich and Schaefer (1994) described frequent examples of poor or lacking communication between different medical professions, or within the surgical team. In a study on severe intraoperative adverse events, Gawande and coworkers (2003) conducted confidential interviews with 38 surgeons asking for factors leading up to the events. Communication breakdown was described as a contributing factor in 43 percent of the events (Gawande, Zinner, Studdert, & Brennan, 2003).

In a different study, the group from Harvard scrutinized 444 surgical malpractice claims with a focus on the role of communication and found communication breakdowns to have contributed to patient harm in 60 (14 percent) of the reviewed cases (Greenberg et al., 2007). The most common factors associated with communication breakdowns were described as status asymmetry (e.g., between resident and staff surgeon), handoffs between OR personnel, and ambiguities of roles, responsibilities and leadership (Greenberg et al., 2007).

In their annual analyses of sentinel events (i.e., severe adverse events that are reported anonymously by U.S. hospitals on a voluntary basis) the Joint Commission for accreditation and safety in healthcare has reported communication as a root cause in over 50 percent of cases from 2004 through 2014 (Joint Commission, 2014).

Prospective observational studies also revealed frequent problems with intraoperative communication. Christian et al. (2006) conducted structured observations during 10 complex general surgical operations, coding safety-influencing system factors and safety-compromising events, as well as potential contributing and compensatory factors. The authors identified communication and information flow, as well as coordination of workload and competing auxiliary tasks as the main safety-influencing system factors (Christian et al., 2006). Notably, Christian and coworkers (2006) observed 88 instances of information loss in
the 10 cases, 86 percent of which disrupted the operative flow. Communication breakdowns and handovers of patient care between providers were found to compromise patient safety in 4 out of 11 instances, while checks and verifications were among the most common compensatory factors that were found to help trap and mitigate errors (Greenberg et al., 2007).

In an effort to characterize communication failures in the OR, Lingard et al. (2004) conducted 90 hours of observations during 48 operations, focusing on communication between team members. Of the recorded 421 procedurally relevant communication events 129 (31 percent) were classified by the authors as communication failures (Lingard et al., 2004). From the observers’ field notes, Lingard and coworkers (2004) characterized communication failures, describing four distinct failure modes: occasion (46 percent); content (36 percent); purpose (24 percent); and audience (21 percent). Occasion failures were described as exchanges in the wrong situation or context, while content failures occurred when inaccurate or insufficient information was transmitted (Lingard et al., 2004). Failures in purpose were observed when the purpose of a communication was unclear or not achieved through the exchange of information (Lingard et al., 2004). Audience failures were defined by Lingard et al. (2004) as communication events in which key actors (either senders or recipients of information) were missing. One third of the observed communication failures were judged by the authors to have the potential to compromise patient safety, for instance by creating inefficiency, tension among team members, delays, and procedural errors (Lingard et al., 2004).

Leadership has been identified as an essential element of teamwork, as conceptualized by Salas et al. (2005). For the purpose of this review I will discuss leadership as an independent nontechnical skill category from the perspective of the surgeon in the OR. Leadership has been described as “the process of influencing others to understand and agree about what needs to be done and how to do it, and the process of facilitating individual and collective efforts to accomplish shared objectives” (p. 8) (Yukl, 2006). A wide and confusing variety of conceptualizations and behavioural constructs of leadership exist, and attributes associated with leadership in surgery cover the entire spectrum of technical and nontechnical skills (Patel et al., 2010). For the purpose of training and assessment, however, leadership
needs to be clearly defined in order to determine learning objectives. In the present thesis, I will refer to leadership as a nontechnical skill of surgeons in the OR, which is distinct from other types such as organizational leadership. Salas et al. (2005) emphasize three main responsibilities of a team leader: to establish and maintain shared mental models among team members; to co-ordinate team behaviours and interactions; and to set expectations about interactions, behaviours and overall performance. Gary Yukl (2012) recently proposed a hierarchical taxonomy to classify leadership behaviours in organizations, distinguishing between task-oriented and relations-oriented behaviours, among others. Examples of task-oriented behaviours include: planning of team activities and assigning tasks; clarifying roles, responsibilities, and objectives; monitoring to ensure team members perform assigned tasks and to identify problems; and problem solving (Yukl, 2012). Relations-oriented behaviours, according to Yukl (2012), include: supporting (e.g., by showing concerns for needs and feelings of team members); and empowering (e.g., involving subordinates in decisions, asking for their input). Surgeon leadership in the OR is, in my opinion, different from individual leadership in organizations as described by Yukl (2012), in that frequently more than one leader is present. The anaesthesiologist, for instance, is often as much a leader as the surgeon, depending on the situation. An example is the induction and intubation phase, during which the surgical sub-team remains relatively passive. Yule et al. (2006) defined the skill category leadership in the NOTSS framework with the skill elements “setting and maintaining standards”, “supporting others”, and “coping with pressure”. Henrickson Parker et al. (2013) recently proposed a comprehensive taxonomy of surgical leadership skills in the OR comprising the elements “maintaining standards”, “making decisions”, “managing resources”, “directing”, “training”, “supporting others”, “communicating”, and “coping with pressure”. One criticism of this “Surgical Leadership Inventory” is that it essentially represents a comprehensive list of nontechnical skills that are relevant for a surgeon in the OR, rather than focusing specifically on skills that are unique to the team leader (Henrickson Parker, Flin, McKinley, & Yule, 2013). Leadership has been associated with safety and performance in healthcare. Gaba (1994) identified poor leadership as a frequent cause for poor team performance in simulated OR emergency situations, highlighting specifically the failure to assume authority in a crisis. Similarly, Helmreich and Schaefer (1994) expressed concerns regarding the lack of leadership they observed in OR teams.
1.2.1.1.3 Personal behaviours

In this category of skills I will discuss personal behaviours and attitudes pertaining to professionalism. Professionalism is a construct of attitudes and behaviours that are strongly associated with the public notion of physicians’ and surgeons’ virtues. The professional demeanour of surgeons has been scrutinized in previous studies. Papadakis et al. (2005) analyzed the cases of 235 physicians that had been disciplined by state medical boards in the U.S. between 1990 and 2003, and identified unprofessional behaviour to be the cause for disciplinary action in at least 74 percent of the cases. As examples for unprofessional behaviour in the examined cases the authors listed the use of drugs or alcohol, unprofessional conduct, conviction of a crime, and negligence, among others (Papadakis et al., 2005). The American Board of Medical Specialties (ABMS) Standing Committee on Ethics and Professionalism reviewed the literature for definitions of professionalism (Wynia, Papadakis, Sullivan, & Hafferty, 2014). The ABMS essentially defined medical professionalism as a belief system rather than a list of behaviours and skills, with the purpose of ensuring that physicians are worthy of the trust bestowed upon them by the public (Wynia et al., 2014). For the purpose of education and assessment, however, a taxonomy of skills, behaviours and attributes is needed to guide educators in ensuring trainees are developing into professionals that are worthy of their patients’ trust. The Accreditation Council for Graduate Medical Education (ACGME) describes professionalism with the following attributes, skills and behaviours: “(1) compassion, integrity, and respect for others; (2) responsiveness to patient needs that supersedes self-interest; (3) respect for patient privacy and autonomy; (4) accountability to patients, society and the profession; and (5) sensitivity and responsiveness to a diverse patient” (ACGME, 2013). The Members of the Medical Professionalism Project placed the patient in the center of their professionalism charter, stating as the main principles of professionalism the dedication to the welfare and autonomy of the patient, as well as the principle of social justice (Medical Professionalism Project, 2002). It should be noted that professionalism has not been included in traditional frameworks of nontechnical skills in aviation (Flin et al., 2003) and health care (Flin & Patey, 2011; Yule, Flin, Paterson-Brown, Maran, & Rowley, 2006). In the NOTSS
framework, some aspect of surgeons’ professionalism were reflected in the element “setting and maintaining standards” pertaining to the skill category “leadership” (Yule, Flin, Paterson-Brown, Maran, et al., 2006). In a recent Delphi consensus on the definition of nontechnical skills for healthcare education, Gordon et al. (2014) included the professional attributes compassion, honesty and integrity in the nontechnical skills category “personal behaviours”. The representation of professionalism as a distinct skill category has clear advantages over the inclusion of related behaviours in other skill constructs, as it facilitates the definition of learning objectives as well as the assessment of learning.

1.2.2 Training Interventions and Initiatives

Around the start of the new millennium, following the publication of “To Err is human”, policymakers, insurers, human factors researchers and healthcare providers worldwide began working on strategies, programs and initiatives to reduce error in medicine (Kohn LT, 2000). A strategy that was recommended by the Institute of Medicine to reduce error and enhance safety in healthcare was CRM training. As a result, CRM principles were among the most commonly implemented strategies used in early programs to improve patient safety and teamwork in healthcare. The late Professor Robert Helmreich, who had been among the pioneers of human factors research in aviation, pointed out the similarities between the OR and the flight deck of an airplane (Helmreich & Schaefer, 1994). In observations of OR teams in a Swiss university hospital Helmreich and Schaefer (1994) identified several behaviours and failures in interpersonal interactions that showed similarities with behaviours previously observed in aviation crews, including communication breakdown, insufficient monitoring of the patient, failure to inform other team members of problems, deficiencies in checklist completion, and poor management of conflicts. Their observations led the authors to the recommendation of implementing CRM principles in surgery, and to train CRM skills of OR teams in simulators (Helmreich & Schaefer, 1994).
1.2.2.1 CRM- and team-training interventions

One of the first CRM programs was implemented by the Vanderbilt University Medical Center in Nashville, Tennessee who hired aviation CRM trainers to administer eight-hour courses to 489 employees from multiple areas of care including the emergency department, trauma unit and the OR (Grogan et al., 2004). CRM principles taught in the course included managing fatigue, creating and managing a team, recognizing adverse events by “red flags”, and cross-checking as a means to improve communication (Grogan et al., 2004). Training was assessed with an end-of-course critique and a survey of attitudes towards CRM skills, and revealed very positive reactions from participants, as well as a significant shift to more positive attitudes about CRM principles and skills (Grogan et al., 2004). In a follow-up study during two subsequent years after CRM training, the group observed cardiac and neurosurgery teams during 30 cases in the OR (France, Leming-Lee, Jackson, Feistritzer, & Higgins, 2008). Although initial participation rates in the training had been high, compliance of OR teams with safety and CRM-principles taught in the courses was below 60 percent (France et al., 2008).

In the U.K., Bleakely et al. (2006) devised a training intervention based on CRM principles and nontechnical skills and administered the training to OR teams of one OR complex at the hospital. At the same hospital, OR teams from a different OR complex were not trained and served as a control group (Bleakley, Boyden, Hobbs, Walsh, & Allard, 2006). The intervention was conducted as a two-day course for local champions and sceptics, which was followed by a one-day seminar on nontechnical skills and patient safety for all OR staff in the training group. After the seminars, preoperative briefings, postoperative debriefings and a near-miss reporting system were introduced hospital-wide (Bleakley et al., 2006). The Safety Attitudes Questionnaire (SAQ), a survey instrument that assesses attitudes of healthcare providers on the six domains teamwork climate, safety climate, perceptions of management, job satisfaction, working conditions, and stress recognition (Sexton, Helmreich, et al., 2006), was administered at baseline (BL) and after one year and revealed a significant shift towards positive attitudes about safety and teamwork in the training group, while the control group showed no change in attitudes (Bleakley et al., 2006). The OR teams
in the control group were subsequently trained as well, and three years following the training the authors surveyed all OR staff regarding the compliance with briefings in the OR (Allard et al., 2007). Of all 118 respondents, 50 percent reported that briefings were never conducted, while 32 percent and 18 percent thought it occurred occasionally or always, respectively (Allard et al., 2007). Interestingly, there was a discrepancy between specialties regarding the perceived frequency of briefings, with 84 percent of surgeons reporting to brief always or at least occasionally, while just over 50 percent of anaesthesiologists and nurses felt accordingly (Allard et al., 2007). Of note, the low response rate to the survey of 44 percent may have biased the results by selecting individuals who felt positively about the intervention.

McCulloch and colleagues (2009) offered CRM training to OR staff at one UK teaching hospital, comprising of a nine-hour course on nontechnical skills and a three-month period of twice weekly coaching in the OR by CRM experts. They demonstrated a significant improvement in attitudes about teamwork and safety climate on the SAQ when comparing baseline (BL) measures with results after the three-month coaching period (McCulloch et al., 2009). Moreover, human-factors experts observed OR teams during 48 procedures at BL and 55 procedures after the coaching period and rated nontechnical skills on the Oxford NOTECHS rating scale, showing a small but significant increase in the teams’ NOTECHS scores after training (McCulloch et al., 2009). Looking at index procedures (laparoscopic cholecystectomy and carotid endarterectomy), the observers also recorded technical errors of the surgical teams by means of the Observational Clinical Human Reliability Assessment (OCHRA), showing a significant decline in error rates from 1.73 at BL to 0.98 after training (McCulloch et al., 2009). This impact on technical performance was particularly interesting. It must be noted, however, that observers were not blinded to the training condition, introducing a potential observer bias. Moreover, improvement in NOTECHS scores was only seen in general surgical teams, while vascular teams did not change. The authors discussed resistance to the training from a few senior members of the vascular team as a possible cause (McCulloch et al., 2009).

Mishra et al. (2009) observed OR teams during 65 laparoscopic cholecystectomies, 26 before and 39 after a training intervention, and rated nontechnical performance of OR teams
on the Oxford NOTECHS scale. The training comprised a 9.5-hour course on nontechnical skills (lectures and exercises aimed at error management, communication techniques and situation awareness) and a three-month period of on-the-job coaching in preoperative briefings (Mishra, Catchpole, & McCulloch, 2009). The authors found a significant improvement in NOTECHS scores of the OR teams at the end of the training intervention, when compared with the BL (Mishra et al., 2009).

In the U.S., Halverson et al. (2009) administered team training to all staff of the North-Western Memorial Hospital in mandatory four-hour courses over a period of 4 weeks. Training was based on CRM principles and comprised lectures, videos, case vignettes and interactive communication exercises (Halverson et al., 2009). The content was structured around team function and communication skills, covering leadership, mutual performance monitoring, shared mental models, the function of preoperative briefings and postoperative debriefings, and specific communication and handoff techniques (e.g., callouts, closed-loop communication) (Halverson et al., 2009). Direct observations in the OR were conducted in the three months before and at six months after the training intervention, showing a 66 percent compliance with the briefing/debriefing at six months, down from 86 percent immediately after the training (Halverson et al., 2009). Observers also recorded communication events and classified communication failures using the classification published by Lingard et al. (2004) (Halverson et al., 2011). Before team training, Halverson et al. (2011) recorded 56 communication failures during 76 hours of observation, reporting inefficiencies, delays and tension among team members as a result of these failures. After team training, 20 communication failures were observed in 74 hours, representing a significant decrease (Halverson et al., 2011). The major limitation of this study again was the lack of blinding of the observers resulting in a potential bias.

The first large-scale multi-site initiatives were led by the Department of Defense (DoD) and the Department of Veterans Affairs (VA) in the U.S., who were part of the Quality Interagency Coordination Task Force founded in 1998 by President Bill Clinton to improve the quality of healthcare (Alonso et al., 2006; Dunn et al., 2007). The Medical Team Training (MTT) program was developed by the VA based on aviation CRM principles, and was first piloted in 2003 in 6 VA facilities (Dunn et al., 2007). Participating hospitals were
offered a choice of program options such as preoperative briefings and postoperative
debriefings, or standardized patient hand-offs, to be implemented in their hospital. CRM
principles were taught in training sessions using practical tools and strategies: SBAR, the
situation-background-assessment-recommendation framework for the standardized hand-off
of patients between providers; rules of conduct for respectful communication between
providers; the concept of “feeling the pinch” as a metaphor for an inner voice that something
is wrong; call out, or speaking up when feeling something is wrong; step back, a strategy to
stop a process and reassess the situation if something appears to be wrong; repeat back and
read back as strategies to ensure effective transfer and understanding of important
information (see also, close-loop communication in the previous paragraph); and the concept
of situation awareness in dynamic environments, among others (Dunn et al., 2007).
Implementation of MTT at each site was conducted in four steps: first, a local
implementation team was selected and trained during a preparation and planning phase;
second, the interactive training sessions were administered to all hospital staff; third, the
program (e.g. preoperative briefings) was implemented and followed-up for several months
by the National Center for Patient Safety; fourth, the MTT program was evaluated (Dunn et
al., 2007). Evaluation comprised the SAQ at BL and post training, self-reported data
regarding safety-relevant behaviours such as briefing, and patient outcomes as assessed by
the VA quality data bank (Dunn et al., 2007). Following the pilot phase, the MTT program
was successively deployed on a large scale and implemented in all VA hospital sites. Wolf
and colleagues (2010) reported their findings from the implementation of MTT in one VA
hospital over the first two years. Attitudes of OR staff improved on all domains of the SAQ,
significantly so for the domains perceptions of management and working conditions (Wolf,
Way, & Stewart, 2010). Compliance rates with briefing/debriefings were high, ranging from
95 to 100 percent, and comparison of 4863 completed briefing/debriefing checklists with
entries in electronic records on 7573 cases from before the implementation showed a
significant decrease in the proportion of cases with delays, from 32 percent before to 19
percent after MTT (Wolf et al., 2010). Moreover, the appropriate and timely administration
of preoperative prophylactic antibiotics as a marker for quality improved from 85 percent in
the first three months post implementation to 97 percent after two years (Wolf et al., 2010).
Paull and colleagues (2010) evaluated compliance with preoperative antibiotic and deep vein thrombosis (DVT) prophylaxis in 74 VA hospitals, comparing compliance rates before and after MTT and implementation of the briefing/debriefing checklist. The 74 facilities reported high implementation-rates of checklist-guided briefings and debriefings between 92 and 98 percent in the first year (Paull et al., 2010). Reviews of charts from all 74 facilities revealed compliance rates with preoperative antibiotic as well as DVT prophylaxis that were significantly higher in the 12 month period after MTT and checklist implementation than during with the 12 month period before (Paull et al., 2010).

In a different study aimed at patient outcome, Neily and coworkers (2010) analyzed electronic records of 182,409 surgical patients from 108 VA facilities, comparing annual mortality rates in the year after implementation with rates in the year before. In the 74 facilities that had undergone MTT, the annual surgical mortality rate had decreased significantly by 18 percent. In the same time period, mortality in the 34 facilities that had not yet undergone training had decreased by only 7 percent, which was non-significant (Neily et al., 2010). The same group evaluated risk-adjusted surgical morbidity in a retrospective chart analysis of 119,393 operative procedures from 74 VA hospitals, of which 42 were in the MTT program and 32 had not yet participated in the program (Young-Xu et al., 2011). The authors demonstrated a significant 17 percent decrease in surgical morbidity in the 42 MTT training facilities, while a non-significant decrease of 6 percent was found in the 32 control facilities (Young-Xu et al., 2011). It should be noted, however, that in all VA studies reporting implementation of MTT formal preoperative briefings and postoperative debriefings were introduced concurrent with the CRM training interventions (Dunn et al., 2007). Since preoperative briefing using a checklist has been shown to independently reduce mortality and complication rates, as demonstrated by large scale studies evaluating the World Health Organization surgical safety checklist (Haugen et al., 2014; Haynes et al., 2009; van Klei et al., 2012), this may have confounded the results with regards to the impact of CRM training.

Another example for a large-scale initiative is the Team Strategies and Tools to Enhance Performance and Patient Safety (TeamSTEPPS) program, which was developed by the DoD and the Agency for Healthcare Research and Quality based on CRM principles and the
teamwork literature (Alonso et al., 2006; Salas, DiazGranados, Weaver, & King, 2008). TeamSTEPPS comprises four modules: leadership; situation monitoring; mutual support; and communication, which are taught in an interactive fashion using practical exercises (Alonso et al., 2006). The TeamSTEPPS web page (http://teamstepps.ahrq.gov) provides an extensive database of readily available tools, videos, webinars and presentations that aid in the design of an individual program. The tools and procedures that are taught in TeamSTEPPS to enhance teamwork are very similar to the CRM concepts of MTT and include the previously discussed communication strategies (e.g., SBAR, check-back, call out, hand-off, closed-loop communication, briefing and debriefing), as well as team concepts such as cross-monitoring (monitoring of other team members’ performance) and mutual support (Weaver, Rosen, et al., 2010).

1.2.2.2 Simulation-based training

Since the beginning of the new millennium, simulation-based training in simulated OR environments has been increasingly used to improve teamwork, communication and other nontechnical skills of surgical teams. The use of simulation to replicate the workplace and allow for safe training of routine and crisis situations in the OR was pioneered in the mid-1980s by David Gaba (1988) who developed a simulated environment comprising of a realistic OR, monitors, equipment, actors in the roles of OR personnel, as well as a human patient simulator with physiologic functions and vital signs that could be controlled by the experimenter (Gaba & DeAnda, 1988). The human patient mannequin allowed for the simulation of critical incidents and intraoperative complications in a realistic setting resembling the real OR (Gaba & DeAnda, 1988). Gaba and DeAnda (1989) conducted an observational study of anaesthesiology residents in a high-fidelity simulation environment to investigate inter-individual differences in the detection and management of the crises, problem-solving strategies, and the effects of errors and deviations (Gaba & DeAnda, 1989). Significant inter-individual differences in the detection and correction of problems were identified between residents that could not be explained entirely by the level of experience, thus leading Gaba and DeAnda (1989) to the conclusion that the management of such problems should be formally taught in residency training. The group subsequently
developed their version of CRM, Anaesthesia Crisis Resource Management, combining the classroom-based CRM training with practice in OR crisis simulations (Howard et al., 1992). Helmreich and Schaefer (1994) similarly conducted simulation-based CRM training interventions with multidisciplinary OR teams, drawing on experience in aviation CRM training.

It was not before the landmark Institute of Medicine report (Kohn LT, 2000) on human error in healthcare, however, that simulation-based training of nontechnical skills gained momentum in the surgical specialties. The group at Imperial College in London, U.K. were among the first to use the simulated OR environment for the training and assessment of surgeons, with regards to both technical and nontechnical skills (Moorthy, Munz, Adams, Pandey, & Darzi, 2005). High-fidelity simulations using a full OR team, a human patient simulator (SimMan, Laerdal, UK) and a synthetic model of a saphenofemoral junction allowed for the completion of entire operations in a realistic environment, while trainers were able to introduce crisis situations and provide feedback on observed performance after the simulations (Moorthy et al., 2005; Moorthy, Munz, Forrest, et al., 2006).

Undre et al. (2007) subsequently used the simulated OR to train entire OR teams in crisis scenarios involving airway problems, intraoperative haemorrhage, or cardiac arrests (Undre, Koutantji, et al., 2007). Technical and nontechnical performance of all involved disciplines was assessed through direct observation by psychologists who discussed participants’ performance after the simulations in debriefings (Undre, Koutantji, et al., 2007). Evaluation of a post-course critique showed that all participants rated this type of training favourably (Undre, Koutantji, et al., 2007). Multiple studies have subsequently replicated the findings, demonstrating feasibility, face validity and positive responses to high fidelity OR simulations as a method to train and assess surgeons’ nontechnical skills in various surgical subspecialties including general surgery (J. Paige et al., 2007; Powers et al., 2008), vascular surgery (Black, Nestel, Kneebone, & Wolfe, 2010), and urology (Lee, Mucksavage, Canales, McDougall, & Lin, 2012).

Paige et al. (2009) used high-fidelity OR simulation with crisis scenarios to teach teamwork competencies to multidisciplinary OR teams in structured debriefing sessions. CRM
principles taught in the debriefing sessions included shared mental models, open communication, role clarity, resource management, cross-monitoring, and situation awareness (J. T. Paige, Kozmenko, Yang, Gururaja, et al., 2009). The authors assessed self-efficacy beliefs of 38 OR staff at BL and after the simulation-based training using a 15-item questionnaire, and showed a significant increase in perceived role clarity, anticipatory response, cross monitoring, and team cohesion (J. T. Paige, Kozmenko, Yang, Paragi Gururaja, et al., 2009). In a follow-up project, additional repetitive simulation-based training and debriefing sessions were administered after the first simulation, resulting in a significant improvement in nine out of the 15 self-assessed competencies in the questionnaire (J. T. Paige, Kozmenko, Yang, Gururaja, et al., 2009). A weakness of Paige’s studies was the reliance on self-assessment, which has been shown to be inherently unreliable in studies involving physicians (Davis et al., 2006).

An important aspect of simulation-based training is the debriefing. Debriefing is led by a designated debriefer and allows trainees to reflect on their performance in the simulation, discuss tasks completed well and identify areas for improvement (Fanning & Gaba, 2007). The reflection on one’s performance in the debriefing is thought to be critical for learning to occur, as will be discussed in the paragraph on experiential learning. The debriefer can either be a trainer or clinical supervisor or, in team simulations, a designated member of the team. Boet et al. (2013) reported results of a randomized controlled trial comparing facilitator-led debriefing to within-team self-debriefing using video recordings of the simulations. The authors enrolled a total of 120 subjects and conducted 40 high-fidelity OR crisis simulations with multidisciplinary teams, which were randomized to either the facilitator-led debriefing or self-debriefing (Boet et al., 2013). Each of the 40 OR teams completed a simulated scenario, underwent the debriefing session, and immediately went on to a different post-training simulation. When evaluating the crisis management performance of the teams through blinded assessment of the video recordings the authors showed a significant improvement in CRM skills in all OR teams from the BL to the post-training simulation, with no difference between groups (Boet et al., 2013).

In summary, multiple studies, interventions and programs have addressed nontechnical performance of OR teams, using a variety of strategies including interactive seminars, on-
the-job coaching, realistic OR simulations, or combinations thereof. Of note, the articles reviewed in this paragraph are limited to studies and interventions targeting surgical teams and/or OR staff, as interventions specifically aimed at surgical trainees will be reviewed separately at a later stage in the present chapter. Interventions range from single hospital, one-time courses to large-scale programs that were implemented in an entire hospital network or health system. Positive results were reported from several studies, including a reduction in communication failures (Halverson et al., 2011), improved compliance with safety-relevant standard procedures (e.g., Wolf et al., 2010; Paull et al., 2010), improvements in observed nontechnical skills (e.g., Mishra et al., 2009; McCulloch et al., 2009), a shift towards more positive attitudes about teamwork and patient safety (e.g., Paige et al., 2009), and even a decrease in the morbidity and mortality of surgical patients (Neily et al., 2010; Young-Xu et al., 2011). Despite weaknesses in the design of several studies, the current evidence supports the training and education of OR teams in nontechnical skills and CRM principles to improve patient safety and efficiency.

1.2.3 Nontechnical Skills in Postgraduate Surgical Education

In the following paragraph I will examine questions and problems pertaining to the teaching and assessment of nontechnical skills in surgical residency training. Before I delve into instructional strategies and methods for assessment, however, I am going to examine the educational environment in order to identify factors that may support or impede the learning of nontechnical skills.

1.2.3.1 How do surgical residents learn nontechnical skills?

Before addressing the question as to how residents learn the nontechnical skills that are critical for safety and efficient performance in the OR, I am going to recapitulate the definition of nontechnical skills. As previously discussed, the construct of nontechnical skills comprises a variety of cognitive processes, behaviours, and attitudes that have been shown to improve task work and enhance team functioning in the OR. Two psychological
models may help explain how residents learn nontechnical skills: observational learning; and experiential learning.

Observational learning, or modeling, is a theory of learning that is rooted in behaviourism, a branch of traditional psychology (Ornstein & Hunkins, 2013). Many early behaviourists have based their theories on observations in animals and subsequently tested them in observations of school children. Examples include Thorndike’s law of effect, describing learning as the establishment of a connection between a situation and a response, and Skinner’s theory of operant conditioning (Ornstein & Hunkins, 2013). A more widely recognized theory of learning in humans is Albert Bandura’s theory of observational learning and modeling (Albert Bandura, 1977). Bandura’s early work involved observations in children, who modeled behaviours that were shown to them in videos (e.g., aggressive behaviours) (Albert Bandura, 1977). While learning from a purely behaviourist viewpoint could be achieved merely from passively observing and modeling others’ behaviours, Bandura himself established the role of cognitive factors in the process of observational learning (A. Bandura, Grusec, & Menlove, 1966). Bandura and coworkers (1966) showed that learning was enhanced when children observed actively by coding the observed behaviours (e.g., by verbalizing what they saw), rather than watching passively. Subsequent studies confirmed these findings in adult learners (Kubany & Slogett, 1991).

In an application of observational learning theory to the learning of nontechnical skills in the OR, residents would learn nontechnical behaviours (e.g., communication techniques, interaction with team members) by observing peers or role models and modeling or adopting similar behaviours in subsequent encounters. A problem with the applicability of this theory is, however, that complex interactions in the OR cannot be compared with video clips that are presented to participants in experimental conditions. Moreover, participation in operations usually does not involve any coding of observed behaviours, which has been shown to be critical for observational learning (Kubany & Slogett, 1991).

One study of residents in the OR suggests a contributory role of observational learning in the attainment of nontechnical skills: in field observations in the OR, Lingard et al. (2002) identified patterns of communication between surgeons, nurses, and anaesthesiologists that
resulted in tension among team members. In an analysis of the effects of such team tension on surgical trainees the authors noted that residents either imitated the communicative behaviour of their superiors in subsequent encounters, with resulting tensions, or withdrew from the interaction altogether (Lingard, Reznick, Espin, Regehr, & DeVito, 2002). I will discuss the mimicking or modeling of observed (negative) behaviour in the following paragraph on the hidden curriculum. These isolated observations, however, do not allow a generalization to more complex behaviours, or cognitive skills such as decision making or situation awareness. Residents are adult learners, and as such are self-directed and have an interest in their own learning and education (Knowles, 1990). One can thus not assume that residents learn nontechnical skills merely from the imitation of others’ behaviour in the OR.

A more appropriate and suitable theory to explain resident learning in the OR is the theory of experiential learning. Experiential learning theory, as popularized by David Kolb (1984), has been conceptualized as a process of knowledge creation from experiences. Kolb (1984) defines learning as “the process whereby knowledge is created through the transformation of experience” (p.38) (D. A. Kolb, 1984). Kolb’s (1984) experiential learning theory is based on learning theories proposed by Kurt Lewin, John Dewey and Jean Piaget, who shared the conception of learning as a process rather than an outcome, and emphasized the significance of experiences (Dewey, 1938; Ornstein & Hunkins, 2013). Based on constructivist theory of learning, experiential learning is conceived as an active process of creating knowledge (A. Y. Kolb & Kolb, 2005). Knowledge created from previous experiences changes the understanding and interpretation of new experiences, which in turn changes the way subsequent encounters are perceived and understood (D. A. Kolb, 1984). Kolb (1984) has conceptualized experiential learning as the result of grasping and transforming experiences. This process of grasping and transforming experience is reflected in the experiential learning cycle (Figure 4). At the basis of learning are experiences, thus in order to acquire new knowledge, skills or attitudes, the learner must actively engage in concrete experiences. He or she then needs to observe, and reflect on the experiences in order to understand or grasp their meaning. Based on the observations, the learner creates new theories, which are then implemented and tested through active experimentation in subsequent encounters. The active experimentation then results in new concrete experiences,
which are again analyzed in reflective observation. The cycle thus continues indefinitely, as each new encounter changes the existing understanding (D. A. Kolb, 1984).

Experiential learning theory lends itself equally well to describe learning in OR simulations (Fanning & Gaba, 2007). The simulation of routine or crisis situations in an authentic environment creates experiences that the learner can actively engage in. In the subsequent debriefing session, the learner reflects on his or her performance in the simulation. This can be enhanced though replay of a video recording, allowing for more detailed observations, or through feedback from a trainer. In the next step, abstract conceptualization, the learner generalizes the findings from observation and reflection and plans modifications of behaviour for subsequent encounters. The new behaviour is then implemented in a subsequent simulation, which creates new experiences that are again analyzed in a debriefing session (Fanning & Gaba, 2007).

![Diagram of Experiential Learning Cycle]

**Figure 4: Experiential Learning Cycle.**

In a grounded theory study using focus group discussions with residents, Teunissen and colleagues (2007b) developed a theoretical construct to describe resident learning in the clinical environment. This construct was translated into a framework of learning in the clinical workplace (Teunissen, Scheele, et al., 2007). In analogy to the experiential learning cycle, resident learning in Teunissen et al.’s (2007b) framework starts with the participation in activities. Experiences made in the activities are then interpreted. The resident’s interpretation of an experience is influenced by “codified knowledge” from textbooks, scientific articles, or treatment guidelines, as well as feedback from peers or supervisors (Teunissen, Scheele, et al., 2007). The interpretation of the experience by the learner then turns it into a personal experience, which the resident transforms into personal knowledge by making sense of the information, thus “constructing meaning”. As described above, this construction of meaning is influenced by external factors, such as comments or feedback from a supervisor (Teunissen, Scheele, et al., 2007). The personal knowledge constructed from the experience may then be subjected to active reflection, for instance regarding future applications of the newly acquired knowledge (Teunissen, Scheele, et al., 2007). The framework of resident learning in the clinical environment aligns well with Kolb’s (1984) experiential learning cycle. An important extension to Kolb’s (1984) theories, however, is the consideration of external factors. Feedback from supervisors, or comments and reactions from other residents or OR personnel must be taken into account, as they can influence how or what residents learn.

In a subsequent qualitative study, Teunissen et al. (2007a) conducted interviews with staff specialists regarding resident learning in the OR. Responses from attending doctors essentially confirmed the previously described framework of resident learning. In contrast to residents’ perceptions of on-the-job learning, however, attending doctors considered feedback from supervisors as the predominant external factor influencing learning (Teunissen, Boor, et al., 2007). In the next paragraph, I will elaborate on external factors that may influence resident learning of nontechnical skills in the OR.
1.2.3.2 Factors influencing Innate Nontechnical Skills and Skill Development in Residency Training

As experiential learning forms the basis of resident learning of nontechnical skills in the OR, several factors need to be considered that may influence how experiences made are integrated into the knowledge of the learners. These include: the formal curriculum; the informal curriculum; and the “hidden curriculum” (Hafferty, 1998). Furthermore, personality traits play a role in how experiences are lived and subjectively viewed and thus potentially influence learning in an experiential learning environment, but may also influence innate nontechnical skills.

1.2.3.2.1 Formal curriculum

The formal curriculum comprises all learning activities that are explicitly stated by the residency program, including didactic teaching in lectures and seminars, and practical teaching of technical skills in simulation-based courses. The formal curriculum includes an explicit statement of the learning objectives and how learning is assessed (e.g., in exams, or by in-training evaluations). The learning objectives of the formal curriculum, which are based on resident training requirements stated by regulatory bodies, also guide residents in self-directed study (e.g., reading of textbooks and scientific articles).

Regulatory bodies for certification and accreditation in North America have integrated nontechnical competencies in their frameworks of resident training requirements and physician competencies, which represent the basis of postgraduate surgical curricula. The Royal College of Physicians and Surgeons of Canada (RCPSC) explicitly states resident training objectives for all surgical specialties and subspecialties in a catalogue of subject-specific knowledge, as well as technical and nontechnical competencies that must be acquired in residency training (RCPSC, 2014). The objectives are structured following the CanMEDS 2005 framework of physician roles and competencies (RCPSC, 2005). Effective and accurate communication is described in the competency “communicator”, though with a strong emphasis on doctor-patient communication (RCPSC, 2005). Communication in the
OR is more specifically addressed in the competency “collaborator”, pointing out the importance of effective communication with members of the OR team. The “collaborator” competency includes further important behaviours pertaining to teamwork. The knowledge of team members’ roles and responsibilities, as well as resource requirements within a healthcare team reflects aspects of situation awareness, as well as the important teamwork mechanism of the shared mental model. Mechanisms of information sharing are further endorsed through the explicit description of the surgical pause in the beginning of an operation and the safe handoff of patients postoperatively (RCPSC, 2010b). The role of the surgeon as a leader is also addressed in the training requirements, specifically the ability to assume the leadership role when required in a team, and to resolve conflicts when necessary (“collaborator” role) (RCPSC, 2010b). Professionalism of the surgeon is represented as a distinct role, including attributes such as honesty, integrity, respect and altruism, and behaviours such as accountability and commitment to patients (RCPSC, 2010b). Although some differences can be found in the wording of the description of behaviours and skill elements between surgical specialties, the nontechnical skills addressed by the RCPSC are essentially identical across surgical specialties and include communication, teamwork, leadership, and professionalism. A limitation of the RCPSC training objectives is the lack of recommendations as to how relevant nontechnical competencies should be taught or assessed.

In the U.S., the ACGME has defined common program requirements describing the organization of training sites and institutions, the educational program, evaluation, and duty hours of residents (ACGME, 2013). The ACGME also provides written program requirements for each of the surgical specialties and subspecialties. Similar to the Canadian training objectives, the program requirements describe nontechnical aspects of performance under the core competencies “interpersonal and communication skills” and “professionalism” (ACGME, 2013). The program requirements state that “Residents must demonstrate interpersonal and communication skills that result in the effective exchange of information and collaboration with patients, their families, and health professionals” (IV.A.5.d) (ACGME, 2013). Examples of behaviours pertaining to the OR are lacking, and the descriptions of behaviours provide only vague representations of doctor-patient communication, communication within a healthcare team and working in a team (ACGME,
A more detailed account of skills and behaviours pertaining to “interpersonal and communication skills” is provided in the milestones that have been defined by the ACGME and the American Board of Surgery (ABS) to facilitate the in-training evaluation of residents (ACGME & ABS, 2014). In the milestones, behaviours and skills are organized as four levels of performance with 1 being the lowest level and 4 representing the performance expected of a resident prior to graduation (ACGME & ABS, 2014). The behaviour of leading the preoperative surgical pause (time-out), for instance, is expected of a resident at level 2. At level 3, residents are required to anticipate needs and requirements in the OR, which could be attributed to situation awareness. Moreover, residents are expected involve team members in problem solving (ACGME & ABS, 2014). At the highest level of performance, residents are expected to assume the role of the leader in a crisis situation.

Professionalism is characterized with attributes and behaviours that are very similar to the Canadian training requirements, including compassion, respect and integrity, altruism, accountability, as well as ethical principles (ACGME, 2013). In contrast to the training objective of Canadian surgeons, the ACGME program requirements explicitly state that programs must conduct formative assessments of residents’ performance in interpersonal and communication skills as well as professionalism in regular intervals (ACGME, 2013). A guide to the evaluation of nontechnical skills is provided in the milestones, though with limited representation of nontechnical aspects of performance (ACGME & ABS, 2014). Specific recommendations as to how nontechnical skills should be taught are not made by the ACGME. The American College of Surgeons (ACS) and the Association of Program Directors in Surgery (APDS) have provided a detailed guide to the training of team-based skills of residents in crisis simulations in the team-based skills module of their Surgery Resident Skills Curriculum, which is available online (ACS/APDS, n.d.).

In the U.K., the Intercollegiate Surgical Curriculum Programme (ISCP) provides a detailed account of nontechnical skills in postgraduate training curricula for the surgical specialties, including a complete syllabus of knowledge, skills and behaviours pertaining to different areas of nontechnical performance that are required of trainees (ISCP, 2013c). Nontechnical skills are represented specifically in the syllabus for “professional behaviours and leadership”, comprising training objectives, knowledge, skills and behaviours pertaining to
communication with patients and colleagues, teamwork, leadership and professionalism, as well as patient safety and time management (ISCP, 2013a). For each individual skill, learning objectives are provided as well as multiple examples and descriptions of knowledge, skills and behaviours pertaining to the skill. In 2013, the General Surgery curriculum was amended with recommendations regarding the use of simulation to develop target skills, ranging from “desirable” to “strongly recommended” (ISCP, 2013a). Simulation is recommended to enhance learning in the nontechnical skill areas “communicator” (e.g., communication with patients and colleagues), “manager” (e.g., teamwork, leadership), and “probity and ethics” (skills and behaviours pertaining to professionalism. Concrete recommendations to the training of nontechnical skills are currently lacking.

The Royal Australasian College of Surgeons (RACS) has developed the Surgical Education and Training (SET) program, comprising of a curriculum framework, general and specific training objectives, and methods for assessment (RACS, 2014). The SET curriculum for General Surgery for instance comprises the competencies communication, collaboration, health advocacy, management and leadership, professionalism and ethics, and scholar and teacher, based on CanMEDS (RCPSC, 2005). Nontechnical performance is assessed on an in-training assessment form with descriptive anchors and exemplar behaviours for different levels of performance in the nontechnical competencies communication, leadership and management, collaboration, and professionalism (Board in General Surgery, n.d.).

In summary, regulatory authorities in North America, the U.K., and Australasia have mandated the integration of nontechnical skills in formal surgical curricula, with differences in the guidance regarding teaching and assessment of the target skills. Communication, teamwork, leadership and professionalism are represented in all training objectives and guidelines reviewed. While nontechnical training objectives in Canada and the U.S. are kept generic with little practical guidance, curricula in the U.K. and Australia provide examples of observable nontechnical behaviours in the OR. The formal curriculum likely has an impact on experiential learning of nontechnical skills in the OR, for instance through targeted nontechnical skills courses. It may also indirectly influence learning through the definition of learning objectives, which may guide residents’ reflection, and staff surgeons’
comments and feedback on behaviours. Definitions of target skills, and provision of exemplar behaviours on assessment forms, may further guide structured formative feedback.

1.2.3.2.2 Informal curriculum

Resident teaching in the informal curriculum includes all teaching activities that are not explicitly stated in the formal curriculum, but are the result of deliberate teaching activities of people in the learning environment. Examples include instructions and feedback that residents receive from supervisors during and after practice in the OR, as well as teaching during ward rounds or in clinics. The quality and quantity of teaching in the informal curriculum strongly depends on the knowledge and experience of the teachers, and even more on their motivation to teach. Learning in the informal curriculum is reflected in Teunissen et al.’s (2007b) framework of resident learning in the workplace, as discussed previously.

1.2.3.2.3 Influence of the Hidden and Null Curriculum

The hidden curriculum is an educational phenomenon that has previously been described and studied in the context of medical education (Gofton & Regehr, 2006). The hidden curriculum comprises expectations, attitudes, values and norms present in an educational environment that influence learners, but are neither explicitly stated in the formal or explicit curriculum nor deliberately taught in the informal curriculum (Hafferty, 1998; Portelli, 1993). Implicit messages that are transmitted in social interactions within and between groups in a learning environment have also been described as constituents of the hidden curriculum (Ornstein & Hunkins, 2013). Examples include attitudes, beliefs and behaviours that are unintentionally conveyed to learners by their teachers (Gofton & Regehr, 2006). The hidden curriculum has a significant impact on learning in an educational environment and may result in unintended (positive or negative) learning outcomes (Portelli, 1993). In medical education, the hidden curriculum has been shown to contribute substantially to the socialization of medical students and residents, particularly with regards to the development
of ethical principles and professionalism (Hafferty & Franks, 1994). Concepts of medical ethics for instance are thought to be acquired primarily through the adoption of values, attitudes and behaviours in the hidden curriculum (Hafferty & Franks, 1994).

Depending on its content, the hidden curriculum can have a positive or negative impact on the formal curriculum. Following Teunissen et al.’s (2007b) conceptualization of resident learning in the OR, the hidden curriculum can act as an external factor that influences the interpretation of experiences and the subsequent construction of knowledge. For instance, a derogatory comment from a staff surgeon about preoperative briefings, although not intended as an educational measure, may influence a resident’s understanding of the relevance of formal team briefings and result in subsequent changes in behaviour (e.g., omission of steps in the briefing). A recent survey of Canadian medical students in clinical rotations revealed a discrepancy between generally accepted professional values, and attitudes and behaviours exhibited by role models in clinical practice (Phillips & Clarke, 2012). Medical students reported observing a variety of negative behaviours and attitudes during their clerkships, including stereotyping, disrespectful comments about patients and discrimination against minorities (Phillips & Clarke, 2012). The hidden curriculum may also influence the development of residents’ nontechnical skills through the adoption of behaviours observed in role models such as staff surgeons, fellows and senior residents (Albert Bandura, 1977). Modeling of poor or dangerous behaviours observed in the OR may result in the adoption of unwanted behaviours by residents, as shown by Lingard et al. (2002) who found residents imitate staff surgeons’ negative communication styles in situations of team tension in the OR.

The hidden curriculum may also impact resident learning and attainment of nontechnical skills through the organizational culture in the educational environment (Hafferty, 1998). Previous surveys in healthcare have demonstrated discrepancies between attitudes of surgeons and other professional groups in the OR regarding principles of CRM, with staff surgeons being less likely to support the principle of flat hierarchies (e.g., junior staff being empowered to challenge decisions by senior staff, and senior staff being open for criticism and suggestions) (Helmreich & Schaefer, 1994; Sexton, Thomas, & Helmreich, 2000). Surgeons have further been found to be less cognizant of personal vulnerability to stress and
As mentioned in the work of Sexton, Makary, et al. (2006), fatigue can significantly impact safety in the operating room (OR). In a survey of surgeons and OR nurses in Scotland, Flin et al. (2006) reported generally positive attitudes about teamwork and safety in the OR, but found staff surgeons to be less supportive of preoperative team briefings than nurses and surgical trainees. Negative attitudes towards safety-relevant behaviours such as briefing, however, may have negative implications for the safety culture of a hospital and by means of the hidden curriculum impede the adoption of desired behaviours by trainees. The importance of hospital safety culture has been underscored in a recent study by Birkmeyer and colleagues (2013) who surveyed 184 OR staff from 22 hospitals regarding attitudes about hospital and OR safety culture and evaluated the relationship of perceived safety culture with patient outcomes through a review of a clinical registry including 24,117 bariatric surgery patients. The authors found a significant association between nurses’ ratings of hospital safety culture and patient outcomes, with rates of serious complications increasing from 1.5 percent in hospitals rated “excellent”, to 2.5 percent and 4.0 percent in hospitals with “very good” and “acceptable” ratings of safety culture, respectively (Birkmeyer et al., 2013). Similarly, surgeons’ perceptions of safety culture in the OR was associated with patient outcome, with a 1.9 percent rate of serious complications among hospitals with an “excellent” OR safety culture, 2.6 percent among hospitals rated “very good”, and 4.2 percent complications among hospitals with an OR safety culture rated only “acceptable” by surgeons (Birkmeyer et al., 2013)

A further concept that must be addressed in this context is the null curriculum. The null curriculum can be described in a simplified way as teaching something by not teaching it (Eisner, 1985). Although the null curriculum represents a distinct phenomenon that is different from the hidden curriculum, it may also impact nontechnical skills education. An example of the null curriculum in a surgical residency program for instance is the absence of dedicated curricular components to teach and assess a particular skill or set of skills, resulting in the notion among residents that these skills are less relevant. As an example, the absence of detailed descriptions of nontechnical aspects of performance in formal surgical curricula may lead staff surgeons and residents to believe that these skills and related behaviours are not important.
The effects of the hidden and null curricula must be recognized within surgical postgraduate education, as they may not only impede the learning and development of desired skills and behaviours, but also result in the attainment of poor and even dangerous behaviours. The integration of nontechnical skills in the formal curriculum, e.g., by dedicating curriculum time to teach these skills, explicitly describing nontechnical skills in resident learning objectives and training requirements, and formally assessing nontechnical skills, may help to alter the content of the hidden and null curriculum by changing the perceived importance of these skills among educators and trainees.

1.2.3.2.4 Personality and Nontechnical Skills

Personality can be defined as “the unique psychological qualities of an individual that influence a variety of behavioural patterns (both overt and covert) across different situations and over time” (“Personality,” n.d.). The relationship of personality with behaviour suggests an influence of personality traits on an individual’s innate nontechnical skills, and may also play a role in the interpretation of activities when learning from experiences in the OR. Based on observations from simulation studies with flight crews, Helmreich (1984) highlighted the role of both personality traits and attitudes on pilots’ flight deck performance. He defined personality and personality traits as “relatively enduring characteristics of the individual that are acquired during development and are resistant to change” (p.583) (Helmreich, 1984). He further postulated: “Attitudes, on the other hand, are less deeply internalized components of the self and are subject to change through a variety of techniques that have been explored and refined by social psychologists, politicians, salesmen, and managers” (p. 583) (Helmreich, 1984). In a discussion of potential implications for selection and training of pilots, Helmreich (1984) saw the role of personality traits, which are not easily changeable, in selection, while identifying attitudes as the target of training interventions. By demonstrating that personality traits and attitudes influenced pilots’ CRM skills relatively independent of each other, he concluded that CRM training would be effective in improving performance in the flight deck (Helmreich, 1984). This was proven in several subsequent studies of pilots and flight crews, although the
authors noted the presence of a certain personality cluster that predicted resistance to such training, as discussed previously (Chidester et al., 1991; Helmreich & Wilhelm, 1991).

Human personality has been the subject of multiple research studies into human behaviour and performance, in the search for implications for selection and training. A variety of personality markers, psychological tools and frameworks have been developed to describe personality. One of the most widely recognized and applied models is the Five-Factor Model (FFM) of human personality (McCrae & John, 1992; Peeters, Van Tuijl, Rutte, & Reynen, 2006). The FFM comprises five traits that describe human personality: Neuroticism (versus: emotional stability); Extraversion; Openness to experience; Agreeableness (versus: antagonism); and Conscientiousness (Costa & McCrae, 2011). I will begin by providing an outline of the “big-five” personality traits and accounts in the literature. Of note, in the present work I will only address normal, i.e., non-pathological, variations of personality traits in humans.

Neuroticism represents traits of human personality that are associated with maladjustment, emotional instability, and a tendency to experience negative affect such as anxiety, anger, sadness, guilt, and shame (Costa & McCrae, 1992). A low tolerance for stress has also been associated with Neuroticism (Costa & McCrae, 1992). Conversely, positive traits in this domain are usually subsumed under the term “emotional stability” (Peeters et al., 2006). Individuals low in Neuroticism are thought to be even-tempered and stable under stressful conditions (Costa & McCrae, 1992).

Extraversion describes an individual’s propensity to be energetic, optimistic, assertive, outgoing, talkative, and to enjoy being around other people (Costa & McCrae, 1992). Although not a distinct domain of personality, introversion represents the lack of Extraversion without necessarily converting positive traits of extraverts into negative ones (Costa & McCrae, 1992). Introverts often do not enjoy being in large groups of people, are quiet and reserved, and less assertive, but not necessarily pessimistic or unfriendly (Costa & McCrae, 1992).

The personality domain Openness to Experience represents traits and attributes such as intellectual curiosity, an active and vivid imagination, open mindedness, and appreciation.
and receptivity for aesthetic (Costa & McCrae, 1992). Openness has been associated with creativity, which may be relevant in decision making (LePine, 2003).

The domain Agreeableness represents traits that pertain to interpersonal attitudes and behaviours such as altruism, sympathy, and cooperation rather than competition (Costa & McCrae, 1992). Agreeable individuals are straightforward and frank, but tend to avoid conflict (Costa & McCrae, 1992). Humbleness and tender-mindedness are further attributes of the agreeable person (Costa & McCrae, 1992).

Finally, Conscientiousness is a personality domain that reflects an individual’s competence and self-control (Costa & McCrae, 1992). The conscientious person is generally hardworking, responsible, organized, task-oriented, and strong-willed (Costa & McCrae, 1992; Peeters et al., 2006).

In analogy to research in aviation, the five-factor model has been used to characterize personality traits of medical students and doctors. In a cross-sectional study of medical students and residents from various surgical and nonsurgical specialties, Hoffman et al. (2010) evaluated participants’ personality profiles on the FFM. Surgery residents were found to score higher in Conscientiousness, Emotional Stability and Extraversion when compared with the general population (Hoffman, Coons, & Kuo, 2010). In a different study, MacNeily and colleagues (2011) assessed personality traits of 86 surgeons and 21 pediatricians using the revised NEO personality inventory (NEO-PI-R), which is a well-tested inventory of normal personality based on the five-factor model (Costa & McCrae, 1992). Compared with non-surgeons, surgeons scored significantly higher on the domain Extraversion (Macneily, Alden, Webber, & Afshar, 2011).

In an attempt to characterize the “surgical personality” McGreevy et al. (2002) evaluated personality profiles of 24 surgical residents using the NEO-PI-R, analyzing scores separately for male and female participants. According to the authors, male surgical residents scored lower on Neuroticism, and higher on Extraversion and Conscientiousness when compared to the general population norm (McGreevy & Wiebe, 2002). Female residents, on the other hand, scored higher on Extraversion, while the other domains were in the average range of the population norm (McGreevy & Wiebe, 2002).
Horwitz and coworkers (2011) assessed personality traits of 65 surgical residents in a cohort study using the NEO-PI-R, showing significantly higher scores on Extraversion, Openness and Conscientiousness among the surgeons when comparing scores with the general population. Agreeableness scores on the other hand were significantly lower among the study sample when compared to the general population (Horwitz, Horwitz, Brunicardi, & Awad, 2011). The authors discussed potential implications for residency training, specifically with regards to the ACGME core competencies. While the authors welcomed the high scores on Conscientiousness with reference to the ACGME competency “professionalism”, they raised concerns over the comparatively low Agreeableness scores (Horwitz et al., 2011). Since specifically the traits Trust, Straightforwardness, and Compliance were low while Altruism was within the population average, Horwitz et al. (2011) interpreted the low Agreeableness as an indicator of deficiencies in communication skills with patients, highlighting the need for targeted training of these skills.

Technical skills, on the other hand, were in the focus of a recent study by Rosenthal et al. (2013), who assessed personality traits of 83 surgical trainees and 12 medial students by means of the NEO-Five Factor Inventory (NEO-FFI), while testing performance on a virtual-reality laparoscopy simulator. When accounting for the confounding factors gender and surgical experience, the authors found no relationship between personality traits and performance on the laparoscopy simulator (Rosenthal et al., 2013). Compared with the general population norm, the cohort analyzed by Rosenthal et al. (2011) scored higher on Extraversion and Conscientiousness, and lower on Neuroticism.

A few studies examined the influence of personality on different aspects of performance. Lievens and coworkers (2002) analyzed personality traits of 631 medical students using the NEO-PI-R, reporting high scores of Extraversion and Agreeableness. When correlating personality scores with year-end exam results, the authors found Conscientiousness to be a significant predictor of academic success (Lievens, Coetsier, De Fruyt, & De Maeseneer, 2002). Conversely, the combination of low Conscientiousness and high Extraversion, specifically for the traits Gregariousness and Pleasure-seeking, was associated with poor academic performance (Lievens et al., 2002). The relationship between team composition in terms of personality and team performance was the topic of a recent meta-analytic review of
the literature (Peeters et al., 2006). Peeters et al. (2006) identified significant correlations of team members’ Agreeableness and Conscientiousness with measures of overall team performance. Conversely, variable composition of teams with regards to members’ Agreeableness and Conscientiousness was associated with poorer performance (Peeters et al., 2006).

In summary, characteristic personality traits appear to distinguish surgeons from the general population, with the most consistent findings across studies being above average scores on Conscientiousness and Extraversion, and normal or lower scores on Neuroticism. Personality traits on the five-factor model seem to correlate with academic performance and some measures of team performance. Personality may determine some attitudes and behaviours that are important for nontechnical aspects of resident performance.

1.2.3.3 Are Nontechnical Skills learned “on the job”?

The evidence collected thus far helps to explain how surgeons learn nontechnical skills. It does not, however, answer a question that is far more important: Do surgical trainees learn nontechnical skills “on the job”?

Moorthy et al. (2005 and 2006) found no difference in non-technical performance of junior, medium-level, and senior surgical trainees in standardized crisis simulations (Moorthy et al., 2005; Moorthy, Munz, Forrest, et al., 2006). Similar results were reported in a different study showing a lack of correlation between training level and non-technical skills of Urology residents in OR crisis simulations (Lee et al., 2012).

Black and colleagues (2010) on the other hand, found significant differences in nontechnical skills between junior and senior surgical trainees, and specialist surgeons, with an ascending grade in total scores on the NOTECHS rating scale in an assessment of 30 vascular surgeons in crisis and non-crisis simulations. A major limitation of the latter study, however, was the fact that the assessments were not conducted in a blinded fashion, introducing a potential observer bias due to the overt age differences between the groups (Black et al., 2010).
Significant correlations of nontechnical performance with the level of specialty training and number of years of training in the U.K. were reported by Crossley and colleagues (2011) who assessed 85 surgical trainees during 404 procedures in the real-life OR. Similar to the previously mentioned studies, however, blinding of observers was limited by the study design (Crossley, Marriott, Purdie, & Beard, 2011). Notably, Crossley and coworkers (2011) found no correlation between NOTSS scores and trainees’ age, or years of training outside of the U.K., suggesting an effect of U.K. residency training on non-technical skill development. These results may be explained by the fact that in the U.K., the training and assessment of non-technical skills is an integral part of the ISCP, as described above, which provides a detailed syllabus of nontechnical skills including training objectives, examples of required skill levels, as well as a guide to assessment techniques (ISCP, 2013b).

The results of these studies show that the current evidence is inconclusive. Thus, the question whether residents learn nontechnical skills to a sufficient degree without targeted training remains unanswered for now.

1.2.3.4 Approaches to the Training of Nontechnical Skills in Surgical Residency

In the previous section I contemplated resident learning of nontechnical skills, including theories of learning and factors influencing the acquisition of nontechnical knowledge, skills and attitudes. In the following paragraph, my focus will be on instructional strategies to address nontechnical skills of surgeons in training. This paragraph was reprinted with permission from Elsevier Inc. from Surgery, 154(5), Dedy NJ, Bonrath EM, Zevin B, and Grantcharov TP. Teaching nontechnical skills in surgical residency: A systematic review of current approaches and outcomes, pages 1000-1008. © 2013 Mosby Inc., with permission from Elsevier.
1.2.3.4.1 Introduction

A growing body of evidence suggests that nontechnical skills (NTS) of surgeons play an important role in patient safety in the operating room (OR). Failures in skills such as communication, teamwork, and leadership have been identified as root causes in approximately 60 percent of major perioperative complications (Christian et al., 2006; Joint Commission, 2014; Greenberg et al., 2007). This is of particular concern, given that an estimated 30 percent of interactions in the OR are deemed communication failures (Lingard et al., 2004). Moreover, patients can be at an increased risk of perioperative complications or death if team members in the OR fail to demonstrate behaviours such as information sharing or briefing (Mazzocco et al., 2009). Introduction of team training interventions has been shown to result in a reduction of communication failures (Halverson et al., 2011), improvements in observed NTS (Awad et al., 2005; K. R. Catchpole, Dale, Hirst, Smith, & Giddings, 2010; McCulloch et al., 2009), and even measurable decreases in surgical morbidity (Young-Xu et al., 2011) and mortality (Neily et al., 2010). As a result, regulatory bodies for accreditation and certification in surgery have recently emphasized the need to teach NTS at a postgraduate level. For example, the Accreditation Council for Graduate Medical Education (ACGME) has included interpersonal and communication skills in the ACGME core competencies, as well as the requirements for accredited general surgery residency programs (ACGME, 2008). To address the need for training in these competencies, multiple resources are available to educators. Examples of such resources include frameworks of NTS, which can be used to define learning objectives and content (Yule, Flin, Maran, Rowley, et al., 2008; Yule, Flin, Paterson-Brown, Maran, et al., 2006), detailed guides for faculty such as the team skills module of the American College of Surgeons / Association of Program Directors in Surgery Surgical Skills Curriculum (ACS/APDS, n.d.), and practical tools to deliver specific training content (Agency for Healthcare Research and Quality, n.d.) Previous reviews have reported on NTS training initiatives across various medical and allied health specialties (D. P. Baker, Gustafson, Beaubien, Salas, & Barach, 2005; Buljac-Samardzic, Dekker-van Doorn, van Wijngaarden, & van Wijk, 2010; Chakraborti, Boonyasai, Wright, & Kern, 2008; Gordon, Darbyshire, & Baker, 2012; McCulloch, Rathbone, & Catchpole, 2011); however, there is little evidence for the effectiveness of NTS training in surgical residency. The objectives of the present
review were to identify interventions that specifically target surgeons in training, determine the main outcomes and strength of evidence for each intervention, and to serve as a reference for the design of future evidence-based curricula.

1.2.3.4.2 Methods

Search strategy: A systematic literature search was conducted using the databases Ovid MEDLINE (1992 to September week 4, 2012), EMBASE (1992 to September week 4, 2012), and PsycINFO (1992 to October week 1, 2012). One author and a full-time librarian independently conducted the search. The medical subject headings: “Leadership”, “Patient safety”, “Communication”, “Cooperative behavior”, as well as the search terms “non-technical skills”, “nontechnical skills”, “crew resource management”, “crisis resource management”, and “teamwork” were linked with the medical subject headings terms “specialties, surgical”, “colorectal surgery”, “general surgery”, “gynaecology”, “neurosurgery”, “orthopaedics”, “surgery, plastic”, “thoracic surgery”, “traumatology” or “urology” using the Boolean operator AND. At this stage, no restrictions were applied to retrieve a comprehensive list of potentially includable articles. In addition to the computerized search, a hand search of the Pubmed and Scopus databases was conducted using references from recent literature reviews on the topic (Hull et al., 2012; McCulloch et al., 2011; O’Connor et al., 2008; Yule, Flin, Paterson-Brown, & Maran, 2006).

Definition of NTS: For the purpose of the literature search and data extraction we defined NTS as “cognitive and social skills that complement technical skills and contribute to safe and efficient task performance” (p. 4) (Youngson & Flin, 2010). A validated taxonomy of surgeons’ NTS was used to further define individual cognitive and interpersonal skills (Yule, Flin, Maran, Rowley, et al., 2008; Yule, Flin, Paterson-Brown, Maran, et al., 2006).

Inclusion criteria: Studies were included in the review if they involved the training and assessment of nontechnical knowledge, skills and attitudes of residents in surgical specialties and subspecialties, as well as the operative specialty of Gynecology. Only original articles, published in English-language peer-reviewed journals were included.
Moreover, studies were required to have an experimental or quasi-experimental design, with reported data pertaining to the effect of a training intervention on targeted NTS.

**Exclusion criteria:** Articles not representing empirical research, such as opinion papers, letters, and commentaries were excluded. Studies not providing a description of the instructional strategy, learning content, and evaluation methods used in the research were excluded.

**Data analysis and grading the quality of evidence:** Data extraction was conducted in a systematic fashion. All articles included in the review were analyzed for target group, training content, instructional strategy, delivery of training, assessment of NTS, and outcome of the intervention. The quality of the evidence was graded using criteria from the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system (Guyatt, Oxman, Kunz, et al., 2008; Guyatt, Oxman, Vist, et al., 2008).

### Results

The initial search of the literature yielded 2,831 records. After removing duplicates and non-English articles, 2,425 records remained that were screened as described. The hand search did not result in additional records. After applying the aforementioned inclusion and exclusion criteria, 148 abstracts remained for review and 40 were selected for full-text analysis. Of these 40 articles, 23 studies were finally selected for the present review (Table 1). Figure 5 illustrates the search and exclusion algorithm in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses format (Moher, Liberati, Tetzlaff, & Altman, 2009).

**Design of studies and main outcome themes:** Four studies were randomized controlled trials (RCT) that compared an intervention group with either a control group (i.e., no additional treatment) (DaRosa et al., 2008; Cordula M. Wetzel et al., 2011), or an alternative treatment group (Knudson et al., 2008; Moulton et al., 2009). Nineteen studies used an observational pre-post design, comparing the participants’ performance after an intervention with their own baseline. A contemporary comparison group was used in one observational
study (Webb et al., 2009). The relevant study details such as study population, design, and limitations are summarized in Appendix 1. Appendix 2 reports a comprehensive synopsis of the instructional strategies, methods of assessment, and outcomes of the selected studies. In these appendices, we combined 2 studies by Paige et al. (2009 a + b) into 1 entry because one of the studies represented an extension and follow-up of the other. Among the reported outcomes, we identified 5 main themes: Patient-centered communication, teamwork, surgical decision making, coping with stress, as well as patient safety and error reduction (Table 1).

Figure 5: Search and exclusion algorithm in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses format.

These outcome themes were analyzed using the GRADE system (Guyatt, Oxman, Kunz, et al., 2008; Guyatt, Oxman, Vist, et al., 2008) (Appendix 3). Two studies reported on >1 outcome theme and were therefore included in >2 outcome summaries (Larkin et al., 2010;
Cordula M. Wetzel et al., 2011). Results of the categorization of the selected articles following the GRADE classification are shown in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>Number of studies (participants)</th>
<th>Limitations, risk of bias</th>
<th>Inconsistency</th>
<th>Directness</th>
<th>Publication bias</th>
<th>Imprecision</th>
<th>Strength of body of evidence (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient-centered</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>communication skills</td>
<td>1 RCT (n=32) 6 Pre-post (n=198)</td>
<td>Not serious Very serious</td>
<td>N/a Not serious</td>
<td>Direct</td>
<td>Not detected</td>
<td>N/a Not serious</td>
<td>Low (++))</td>
</tr>
<tr>
<td>Teamwork</td>
<td>2 RCT (n=34) 8 pre-post (n=236)</td>
<td>Not serious Serious</td>
<td>Not detected</td>
<td>Direct</td>
<td>Not detected</td>
<td>Not detected</td>
<td>Moderate (+++)</td>
</tr>
<tr>
<td>Surgical decision making</td>
<td>1 RCT (n=63) 2 pre-post (n=81)</td>
<td>Serious</td>
<td>N/a Serious</td>
<td>Direct</td>
<td>Not detected</td>
<td>Not serious</td>
<td>Low (++)</td>
</tr>
<tr>
<td>Coping with stress</td>
<td>1 RCT (n=16) 2 pre-post (n=56)</td>
<td>Not serious Serious</td>
<td>N/a Not detected</td>
<td>Direct</td>
<td>Not detected</td>
<td>N/a Not serious</td>
<td>Low (++)</td>
</tr>
<tr>
<td>Patient safety and error reduction</td>
<td>2 pre-post (n=59)</td>
<td>Serious</td>
<td>Not detected</td>
<td>Direct</td>
<td>Not detected</td>
<td>Not serious</td>
<td>Very low (+)</td>
</tr>
</tbody>
</table>

Table 1: Strength of body of evidence, organized by main outcome themes.

GRADE system: ++++ = high quality of evidence; +++ = moderate quality of evidence; ++ = low quality of evidence; + = very low quality of evidence. RCT = randomized controlled trial.
**Patient-centered communication:** Overall, 7 studies reported outcome pertaining to patient-centered communication skills: 1 RCT (Moulton et al., 2009), and 6 observational studies using a pre-post design (Chandawarkar et al., 2011; Gettman et al., 2008; Klaristenfeld, Harrington, & Miner, 2007; Larkin et al., 2010; Razack et al., 2007; Webb et al., 2009). Four studies represented one-time interventions that were completed in a single day (Chandawarkar et al., 2011; Gettman et al., 2008; Moulton et al., 2009; Razack et al., 2007), whereas 3 studies were part of a larger initiative spanning several weeks (Klaristenfeld et al., 2007; Larkin et al., 2010; Webb et al., 2009) (Appendix 1). The majority of studies combined didactic methods (e.g., lectures, group discussions, video presentations, reading assignments) with a practice session comprising of role-play exercises or standardized patient encounters (Chandawarkar et al., 2011; Klaristenfeld et al., 2007; Larkin et al., 2010; Razack et al., 2007; Webb et al., 2009). Outcomes were: significant improvements in observed case-specific communication (Chandawarkar et al., 2011), empathic responses to patients (Larkin et al., 2010), and residents’ perception of own communication skills (Klaristenfeld et al., 2007; Razack et al., 2007). In contrast to the combined didactic and practice methods, two studies used predominantly simulation-based approaches, composed of challenging patient encounters, followed by structured feedback in debriefing sessions (Gettman et al., 2008; Moulton et al., 2009). After this simulation-based, “exposure-first” approach, residents’ perceived competence in communicating bad news improved significantly (Gettman et al., 2008); similarly, in 1 RCT, observed communication skills were significantly better in the intervention group receiving structured feedback after the patient encounter, than in the control group (no feedback) (Moulton et al., 2009).

**Teamwork:** Overall, 10 studies reported outcome pertaining to teamwork knowledge, skills and attitudes: two RCTs (Knudson et al., 2008; Cordula M. Wetzel et al., 2011) and eight observational studies (Gettman et al., 2009; Hamilton et al., 2012; Koutantji et al., 2008; Larkin et al., 2010; Marr et al., 2012; J. T. Paige, Kozmenko, Yang, Gururaja, et al., 2009; J. T. Paige, Kozmenko, Yang, Paragi Gururaja, et al., 2009; Peckler, Prewett, Campbell, & Brannick, 2012). The studies were diverse in terms of instructional strategies, methods for training and assessment, and outcomes. One half of these studies were one-day interventions with post-test evaluations on the day of the training (Gettman et al., 2009; Koutantji et al., 2008; J. T. Paige, Kozmenko, Yang, Paragi Gururaja, et al., 2009; Peckler et al., 2012;
Cordula M. Wetzel et al., 2011), whereas the other half distributed the training over a period of time ranging from three to five weeks with once-weekly sessions (Hamilton et al., 2012; Knudson et al., 2008), to three-month periods with multiple sessions (Marr et al., 2012; J. T. Paige, Kozmenko, Yang, Gururaja, et al., 2009) and integrated curricula with recurrent sessions over the duration of an academic year (Larkin et al., 2010).

The majority of reported teamwork-related outcomes were positive: Significant improvements in teamwork knowledge (Knudson et al., 2008; Peckler et al., 2012), a significant shift to positive attitudes towards teamwork (Koutantji et al., 2008; J. T. Paige, Kozmenko, Yang, Gururaja, et al., 2009; J. T. Paige, Kozmenko, Yang, Paragi Gururaja, et al., 2009), significant improvement in observed skills and behaviours in simulated ORs (Gettman et al., 2009; Cordula M. Wetzel et al., 2011), trauma bay simulations (Hamilton et al., 2012), and real-life trauma resuscitations (Knudson et al., 2008; Marr et al., 2012). Similar to the interventions focusing on communication, two main instructional strategies were identified: Didactic teaching followed by practice (Koutantji et al., 2008; Larkin et al., 2010; Peckler et al., 2012), and high-fidelity simulation followed by feedback in debriefing sessions (Gettman et al., 2009; Hamilton et al., 2012; Marr et al., 2012; J. T. Paige, Kozmenko, Yang, Gururaja, et al., 2009; Cordula M. Wetzel et al., 2011) (Appendix 2). One RCT compared a didactic approach with simulation-based training and showed significantly better teamwork and crisis management skills in the simulation group as observed in recorded real-life trauma resuscitations (Knudson et al., 2008). Significant improvements in observed teamwork skills in a simulated crisis scenario were also demonstrated in an RCT after an intervention to teach acute stress coping strategies (Cordula M. Wetzel et al., 2011).

Teamwork was trained and assessed in either an OR environment \( (n=6) \) (Gettman et al., 2009; Koutantji et al., 2008; Larkin et al., 2010; J. T. Paige, Kozmenko, Yang, Gururaja, et al., 2009; J. T. Paige, Kozmenko, Yang, Paragi Gururaja, et al., 2009; Cordula M. Wetzel et al., 2011), or in trauma and emergency situations \( (n=4) \) (Hamilton et al., 2012; Knudson et al., 2008; Marr et al., 2012; Peckler et al., 2012). Simulated scenarios included cardiovascular and respiratory events such as intraoperative bradycardia, myocardial infarction, CO\(_2\) embolism, pneumothorax, anaphylaxis, cardiac arrests (Gettman et al., 2009;
Koutantji et al., 2008; J. T. Paige, Kozmenko, Yang, Gururaja, et al., 2009; J. T. Paige, Kozmenko, Yang, Paragi Gururaja, et al., 2009), anesthesiology problems such as malignant hyperthermia and light anesthesia (J. T. Paige, Kozmenko, Yang, Gururaja, et al., 2009), surgical complications such as intraoperative haemorrhage and gallbladder perforation (Koutantji et al., 2008), or stroke during carotid endarterectomy (Cordula M. Wetzel et al., 2011), trauma resuscitations involving pneumothorax and airway problems (Marr et al., 2012), blunt and penetrating abdominal trauma (Marr et al., 2012), multiple trauma and trauma in pediatric and obstetric patients (Knudson et al., 2008), and obstetric emergencies (Pliego, Wehbe-Janek, Rajab, Browning, & Fothergill, 2008).

**Surgical decision making:** Surgical decision making was the training objective and main outcome in three studies: one RCT (DaRosa et al., 2008) and 2 observational studies (Jacklin, Sevdalis, Darzi, & Vincent, 2009; Scott, Hameed, Evans, Simons, & Sidhu, 2008). Decision making was either taught in a generic context of clinical decisions pertaining to perioperative risk (Jacklin et al., 2009) and trauma management (Scott et al., 2008), or in specific settings focusing on intraoperative decisions during laparoscopic cholecystectomy (DaRosa et al., 2008). Instructional strategies included interactive didactic teaching as well as practice sessions (DaRosa et al., 2008; Jacklin et al., 2009; Scott et al., 2008). Outcomes were assessed through knowledge tests (DaRosa et al., 2008; Jacklin et al., 2009; Scott et al., 2008) and by direct observation during a standardized procedure (DaRosa et al., 2008). In an RCT focusing on decision making and error avoidance during laparoscopic cholecystectomy, the intervention group performed significantly better on a knowledge test than the control group, although no difference in technical performance was found between the groups (DaRosa et al., 2008). In the setting of surgical trauma, one intervention resulted in significant improvement in decision making with knowledge retention over 6 months (Scott et al., 2008). Last, in an intervention to improve perioperative risk assessment, no effects on accuracy of judgment and risk assessment were found, although there was a significant improvement in the consistency of the trainees’ judgments (Jacklin et al., 2009).

**Coping with stress:** The ability to cope with stress was trained and assessed in one RCT (Cordula M. Wetzel et al., 2011) and two observational studies (Larkin et al., 2010; Pliego et al., 2008). In these studies, training objectives and outcome were diverse. A human factors
Curriculum teaching stress management and coping strategies resulted in a significant increase in perceived stress levels (Larkin et al., 2010). In contrast, a one-day intervention targeted at acute stress management showed lower stress levels as measured by heart rate variability, and a significant increase in the use of coping strategies during a crisis simulation in the intervention group, whereas no effect was shown in measurements of salivary cortisol levels and a psychological test (Cordula M. Wetzel et al., 2011). Last, a significant improvement in perceived stress-hardiness was documented after a simulation-based training intervention that lacked specific stress management training (Pliego et al., 2008).

**Patient safety and error reduction:** Two studies focused on patient safety (Arora, Sevdalis, et al., 2012) and error reduction (Brannick, Fabri, Zayas-Castro, & Bryant, 2009) in a surgical context. A classroom-based, half-day intervention resulted in significant improvement in a knowledge test and enhanced self-reported awareness of patient safety issues (Arora, Sevdalis, et al., 2012). In an intervention seeking to reduce human factor errors, participants demonstrated significant improvement in one out of three categories of a situational judgment test (Brannick et al., 2009). However, prospectively collected complication reports showed a significant decrease over time in complication and error rates after the intervention (Brannick et al., 2009).

1.2.3.4.4 Discussion

NTS have been recognized as an integral part of surgical core competencies, and accreditation and certification bodies recommend these skills to be taught in surgical residency. However, in view of increasing workload and work-hour limitations, curriculum time is often scarce. Therefore, interventions should be as efficient and effective as possible. In the present systematic review we identified interventions to teach NTS in surgical residency programs. A meta-analytical review was not feasible owing to the variability of training methodologies and outcome measures, as well as the paucity of RCTs.
Instead, we identified 5 main outcomes themes and appraised the strength of evidence for these using a systematic approach.

**Patient-centered communication:** Communication and interaction with patients and their relatives is considered a basic competency of a surgeon and constitutes an integral part of the ACGME program requirements for graduate medical education (ACGME, 2008). Studies identified for the present review used instructional strategies and methods that were largely in agreement with current ACGME recommendations for program directors suggesting the use of interactive teaching methods including role-play, video demonstrations, and group discussions (ACGME, 2011). However, despite predominantly positive reports on the effects of targeted training on communication skills, we found serious limitations that reduced the overall strength of the evidence. First, with the exception of one RCT (Moulton et al., 2009), studies did not use blinded observers to rate communication skills (Chandawarkar et al., 2011; Gettman et al., 2008; Larkin et al., 2010; Webb et al., 2009). Second, some studies relied exclusively on participant self-assessment to evaluate the impact of the training (Gettman et al., 2008; Klaristenfeld et al., 2007; Razack et al., 2007). Last, attendance rates were low in one study (Klaristenfeld et al., 2007) despite the mandatory nature of the intervention, with a potential for selection bias. Therefore, following the structured GRADE approach, we rated the overall strength of the body of evidence as “low”. Nevertheless, current evidence permits us to conclude that basic communication skills can effectively be taught to residents by means of simulated patient encounters in conjunction with structured, formative feedback (Moulton et al., 2009), and that combining patient communication scenarios with basic procedural tasks in standardized modules allows for time-efficient training and assessment of technical and nontechnical skills in surgical curricula (Kneebone et al., 2006; Moulton et al., 2009).

**Teamwork:** Teamwork can be defined as “…an interrelated set of team member thoughts, behaviors, and feelings needed for the team to function as a unit” (p. 1003) (Salas et al., 2008), and was recently described to encompass 5 core components: Team orientation, mutual performance monitoring, backup behaviour, adaptability, and team leadership (Salas et al., 2005). Following the ACGME core competencies, residents are required to demonstrate interpersonal and communication skills, including the ability to “work
effectively as a member or leader of a health care team” (p. 9) (ACGME, 2013). Although the majority of interventions analyzed for the present review showed positive effects on teamwork knowledge, skills and attitudes, limitations must be considered. The majority of the evidence was established in observational studies lacking control groups. With one exception (Hamilton et al., 2012), none of the observational studies used blinded observers to rate teamwork skills, resulting in a high risk of bias. Moreover, inconsistencies were noted in some observational studies with a lack of training effect in a subgroup (Peckler et al., 2012) or the entire sample (Koutantji et al., 2008; Larkin et al., 2010).

The overall strength of the body of evidence supporting the effectiveness of teamwork training in surgical residency was rated as “moderate”. Notably, the most commonly used instructional strategy comprised of high-fidelity crisis simulations followed by debriefing or feedback sessions (Gettman et al., 2009; Hamilton et al., 2012; Marr et al., 2012; J. T. Paige, Kozmenko, Yang, Gururaja, et al., 2009; J. T. Paige, Kozmenko, Yang, Paragi Gururaja, et al., 2009; Peckler et al., 2012). This simulation-based approach was shown to be superior to a purely didactic approach with regards to skill acquisition (Knudson et al., 2008). These findings are in agreement with generally accepted concepts of training: The need for opportunities to practice in realistic environments to enhance generalization and transfer of skills on the job (Baldwin & Ford, 1988; Salas, Rhodenizer, & Bowers, 2000) and the importance of performance assessment to provide structured feedback in debriefing sessions, allowing trainees to reflect on their behaviour and remediate mistakes (Rosen et al., 2008).

Other evidence-based strategies to enhance acquisition and generalization of team-based skills in simulation-based interventions include the use of video examples of positive and negative behaviour and video-based active observation of own and peer performance in debriefing sessions (Kubany & Slogett, 1991); following a “teaching-last” approach by beginning training sessions with an exposure to simulation or role-play, followed by structured debriefing and didactic content (Zendejas, Cook, & Farley, 2010); and distributing multiple training sessions over several weeks or months rather than conducting “massed” training in a single intervention (J. T. Paige, Kozmenko, Yang, Gururaja, et al., 2009).
Despite the moderate level of evidence supporting teamwork training in surgical residency, combining simulation-based training with the aforementioned concepts may allow residents to experience and learn from challenging situations in an environment reflecting their real workplace. This type of intervention can be implemented in most surgical curricula at any stage of residency. However, basic teamwork skills such as closed-loop communication, handover, and briefing do not necessarily require a full-scale OR simulation, but can be trained in role-play exercises or part task trainers (Beaubien & Baker, 2004). The fidelity of these simulations can be enhanced by combining them with procedural tasks (Moulton et al., 2009). Educators can draw on multiple available teaching resources and practical guides to aid in the design of simulation-based NTS curricula (Agency for Healthcare Research and Quality, n.d.; ACS/APDS, n.d.). Equally, reliable rating scales (Yule, Flin, Maran, Rowley, et al., 2008; Yule, Flin, Maran, Youngson, et al., 2008) are available for structured, formative feedback, which can serve as a means for ongoing monitoring of adequate skill development throughout residency training, ensuring that competencies in teamwork and related NTS show adequate progress.

**Surgical decision making:** A widely recognized model discriminates between four types of decision making in high-risk environments: Recognition-primed, rule-based, analytical, and creative (Flin, Youngson, et al., 2007). A recognition-primed decision making approach is thought to be the most efficient and is generally used by experienced surgeons (Gary Klein, 1998), whereas novice surgeons often rely on rule-based decision making by reflecting on their knowledge of the procedure at hand (Flin, Youngson, et al., 2007). Analytical decision making requires more time because it involves the comparison of options and the assessment of risk before a decision is made and implemented (Flin, Youngson, et al., 2007). Last, creative decision making, which involves generating an action plan in an unfamiliar setting plays only a subordinate role in surgery (Flin, Youngson, et al., 2007). Various types of decision making were trained and assessed in the reviewed studies limiting comparability of the outcomes. Overall, the strength of the evidence was “low” owing to high attrition rates (DaRosa et al., 2008; Scott et al., 2008) and inconsistencies in the evidence regarding training effect (Jacklin et al., 2009).
Nevertheless, the evidence suggests that designated training of procedure-related decision making composed of video examples, group discussions, and hands-on practice is likely to result in improved decision making in the targeted procedures, while addressing both rule-based and recognition-primed mechanisms (DaRosa et al., 2008). However, the number of procedures that can be addressed by such targeted training is limited by available curriculum time. Therefore, a feasible approach may be to integrate discussions about decision making into technical components of surgical curricula. Elements of both procedural and crisis-related decision making can be integrated in hybrid simulations such as high-fidelity crisis simulations encompassing virtual reality or bench top models of procedures (Andrew, Plachta, Salud, & Pugh, 2012).

**Coping with stress**: The ability to cope with acute stress in a crisis situation is an important skill for surgeons, which correlates with technical and nontechnical surgical performance (C. M. Wetzel et al., 2010). Targeted stress management training, including individual coping strategies, relaxation techniques, and mental rehearsal, can reduce physiological stress responses and improve performance and teamwork in a simulated crisis situation (Cordula M. Wetzel et al., 2011). Coping with acute stress should, however, be distinguished from the management of chronic stress. For instance, a rating tool for chronic stress pertaining to life-events was used in one observational study, limiting comparability of results with studies targeting acute stress (Larkin et al., 2010). Furthermore, strength of evidence was limited owing to the inherent unreliability of self-assessment and self-report (Davis et al., 2006; Pliego et al., 2008). The overall strength of the evidence was thus graded as “low”. A feasible approach to teaching acute stress management in future curricula could be to integrate this topic in group discussions or debriefing sessions after high-fidelity crisis simulations.

**Patient safety and error reduction**: Patient safety and human error in surgery has become a priority among healthcare providers and patient safety agencies (Kohn LT, 2000). Large-scale initiatives have shown significant effects of surgical safety checklists and preoperative briefings on patient safety (Paull et al., 2010) and perioperative morbidity and mortality (Haynes et al., 2009; van Klei et al., 2012). Patient safety and error management, including tools such as briefings and checklists, constitute elements of Crew Resource Management
training. Elements of Crew Resource Management training have been implemented in multiple healthcare facilities (Dunn et al., 2007) with resulting medium-term reductions in perioperative morbidity (Young-Xu et al., 2011) and mortality (Neily et al., 2010). In the present review, however, the evidence is flawed by a high risk of bias owing to low participation rates (Arora, Sevdalis, et al., 2012; Brannick et al., 2009). Although error and complication rates decreased significantly in one study (Arora, Sevdalis, et al., 2012), this evidence should be interpreted with caution because complication reports and on-the-job errors were prospectively collected by the study participants themselves.

Even though the strength of the evidence supporting error and patient safety training for residents was low, we do believe that it is very important to raise residents’ awareness of human error and patient safety issues. A feasible approach could be including instructions and exercises on the correct use of briefings and checklists in surgical curricula. Because large-scale initiatives have shown a significant benefit of these tools (Haynes et al., 2009; Neily et al., 2010; van Klei et al., 2012), and it has been shown that checklist-guided briefing increases compliance with preoperative antibiotic administration (Lingard et al., 2011), reduces unexpected delays (Nundy et al., 2008), and decreases the number of communication failures per procedure (Lingard et al., 2008), it seems commendable to include these in residency training. Moreover, knowledge of human error and the role of human factors and teamwork in patient safety could be taught in a seminar or group discussion using case vignettes or videos of adverse events in surgery and other high-risk industries (Flin, Yule, et al., 2007).

**Implications for future curricula and research:** Specific training for NTS in residency should be efficient and effective. It should address all relevant aspects of nontechnical performance and allow for ongoing assessment. An evidence-based approach to achieve these goals is the use of simulation-based training. Here, NTS training can be combined with procedural tasks, and performance can be assessed using validated rating systems (Sharma, Mishra, Aggarwal, & Grantcharov, 2011). To optimize the resident’s learning opportunity and continuously improve on these skills throughout the entire duration of residency, we feel that training and assessment of NTS should begin early and content should be adapted to the level of training. Early exposure of residents to NTS training may also minimize the
adoption of negative behaviours from peers and role models through the “hidden curriculum” (Hafferty, 1998). For that reason, training in NTS should not end with the completion of residency, but attending surgeons should also be trained in the assessment of NTS. This approach has been shown to be feasible; however, specific training and the commitment of the attending surgeons are prerequisite for its success (Yule, Flin, Maran, Youngson, et al., 2008). Regular feedback from attending surgeons on residents’ NTS in the OR could help to enhance the impact of curricular training interventions and optimize experiential learning on the job.

Future research in this field should focus on the long-term effects of training interventions in prospective, randomized trials to deliver the still lacking evidence base for competency based training of NTS in surgical residency. Furthermore, applicable benchmarks of NTS need to be determined for each training level in order to ensure appropriate progression of skills throughout the residency.

1.2.3.5 Considerations for the Selection of Instructional Strategies for Nontechnical Skills Education

A number of factors should be considered when selecting appropriate instructional strategies for nontechnical skills education in surgical residency. First, surgeons are adults when they enter residency training. Hence, principles of adult learning should be taken into account. Characteristics of the adult learner have been conceptualized by Malcolm Knowles (1990). According to Knowles (1990), adults are self-directed in their learning, which should be considered in educational activities, for instance by actively involving learners to take responsibility for their education. Moreover, adults need to connect learning to previous experiences. This is reflected in the theory of experiential learning (D. A. Kolb, 1984) and, more specifically concerning residents, in Teunissen and coworkers’ (2007) framework of learning in the clinical workplace. Furthermore, adult learners are goal-oriented and need clearly defined learning objectives (Knowles, 1990). These should be explicitly stated, either in the formal curriculum or at the outset of an educational activity. The learning must have a
direct relevance to the learner’s field of work and should be of practical use to be perceived as important by the adult learner (Knowles, 1990). Finally, external factors that may potentially impede with resident learning should be considered, including hidden and null curricula, differences in innate skills, and individual learning styles. At this point, I will discuss four groups of instructional strategies that are well tried and commonly used in nontechnical skills education and training, both in aviation and healthcare.

**Lectures and small-group seminars:** Lectures are still among the most popular teaching methods in undergraduate medical education. Utilizing a transmission or direct instruction model of teaching, cognitive content can be conveyed in a time efficient way by providing all the essential information and explanations required to understand the subject matter (Kirschner, Sweller, & Clark, 2006). Direct instruction methods such as didactic lectures and small-group seminars are popular among learners and educators in medical education, since large amounts of content can be conveyed in a relatively short time (Haidet, Morgan, O'Malley, Moran, & Richards, 2004). According to accepted concepts of cognitive psychology, learning takes place when information is transferred from short-term to long-term memory, which is accomplished by the working memory (Kirschner, et al., 2006). In long-term memory, an infinite amount of information can be stored and recalled at any time (Ornstein & Hunkins, 2013). The effectiveness of learning is determined by the rate at which information is processed and organized by the working memory (Ornstein & Hunkins, 2013). This depends on multiple factors, including disposition and readiness of the individual learner, presentation and organization of the material, and reinforcement by the teacher, all of which are characteristic of behaviourist theory (Ornstein & Hunkins, 2013). Transmission teaching through didactic lectures and seminars is a characteristic feature of traditional educational philosophy (e.g., essentialism, which is aimed at the acquisition and mastery of essential skills and knowledge) (Ornstein & Hunkins, 2013). Such teaching is selected and organized by the teacher following a prescribed curriculum, and transmitted in a classroom or lecture hall setting, while learners are passive recipients of the information (Ornstein & Hunkins, 2013). Critics of this traditional philosophy and teaching method have pointed out the lack of critical thinking and problem solving on the part of the learners, and the development of a passive attitude towards one’s own education and learning, which is in contrast to the needs of the adult learner (Haidet, et al., 2004). Seminars represent an
alternative to lectures with the opportunity to present information in an efficient way, while allowing for involvement of the learner through interactive components.

**Behaviour or video modeling strategies:** Video modeling is a frequently used instructional approach in CRM- and nontechnical skills training courses (e.g., (Flin, Yule, et al., 2007)). Behaviour modeling is particularly useful to teach interpersonal skills, as it is often more practicable to explain complex skills or behaviours by demonstrating recordings or reproductions of the target skill in the appropriate context than creating a theoretical description. Practicality, however, does not ensure learning of the desired behaviour. Behaviour modeling training is based on Bandura’s (1977) theory of observational learning, which was described earlier in this text. The theoretical principle of behaviour modeling is that the learner observes a particular skill or behaviour, memorizes and reproduces the behaviour, and then applies it in the real environment, e.g., on the job (Baldwin, 1992). Several factors should be considered to ensure effective learning when designing educational activities involving behaviour modeling. First, learners need to memorize the behaviour to apply it in subsequent practice sessions or on the job (Baldwin, 1992). Kubany and Sloggett (1991) have demonstrated that observational learning using video recordings of the target behaviour was significantly enhanced when learners were required to code observed behaviour, i.e. take notes while watching the video. The provision of clear learning points prior to the demonstration of the target behaviour has been shown to enhance learning of procedural knowledge and skills, as demonstrated in a recent meta-analysis on design characteristics of behaviour modeling training (Taylor, Russ-Eft, & Chan, 2005). This finding is in agreement with generally accepted principles of adult learning, as previously mentioned. With regards to the types of behaviours shown, Baldwin (1992) has shown that the demonstration of both positive and negative examples of behaviours and skills enhances the generalization of the skill or behaviour. It should be noted, however that the demonstration of mixed positive and negative behaviour models reduces the simple reproduction of the skill, when compared with demonstrations showing only positive behaviour (Baldwin, 1992). Taylor and coworkers (2005) confirmed Baldwin’s (1992) findings in their meta-analysis, showing mixed (positive and negative examples) models to be superior to positive-only models with regards to the generalization of skills by learners, as well as the transfer of skills to different settings, i.e. application on the job.
Organizational factors may also influence effects of behaviour modeling training. Specifically, training of the trainees’ supervisors has been shown to enhance behaviour changes on the job, possibly through improved feedback and support from superiors (Taylor et al., 2005). In summary the effectiveness of behaviour modeling training can be enhanced by defining learning points, demonstrating both positive and negative examples of the target behaviour and engaging trainees in active rather than passive observation. Behaviour modeling can also enhance learning when used as an adjunct to lectures or seminars (Baldwin, 1992). On the organizational level, it may be beneficial to involve supervisors in the training to ensure behaviours are positively reinforced in the actual workplace. In the setting of surgical residency training, staff surgeons as the direct supervisors of residents may be trained to ensure effective transfer of skills and a continuation of learning in the informal curriculum, i.e., in the OR. Involvement of staff surgeons in nontechnical skills training may also have positive effects on the hidden curriculum and foster an educational environment in which nontechnical skills and behaviours are welcomed and positively reinforced.

**Simulation-based training:** Simulation has gained wide acceptance and popularity as an instructional method to train nontechnical skills in healthcare. A recent meta-analysis of instructional methods used in the education of health professionals has shown that simulation was superior to other methods, particularly when teaching behaviours (Cook et al., 2012). Simulation allows learners to make experiences and to learn from these experiences in an artificial environment or a classroom that resembles the real world (Lederman, 1984). As such, simulation can be described as a practical application of experiential learning theory, as detailed previously (D. A. Kolb, 1984). The reflection on experiences and the creation of abstract ideas to generalize the findings are important cognitive processes in the creation of knowledge from experiences, as discussed in the section on experiential learning. Moreover, new ideas and concepts derived from the experiences gathered in the simulation can be applied in subsequent simulations, reflecting the process of active experimentation (D. A. Kolb, 1984).

In simulation learners reflect on, and make sense of experiences in the debriefing session (Lederman, 1984). Commonly, debriefing is facilitated by a designated debriefer who is
usually also the learner’s teacher or supervisor. In debriefing, the role of the facilitator is not primarily to provide feedback to the learners, but to encourage active reflection on their own performance (Fanning & Gaba, 2007). The facilitator may, however, provide guidance to the reflection and analysis of performance, depending on the complexity of the simulation and previous experience of the learner in the particular learning environment (Fanning & Gaba, 2007). The level of facilitation required also depends on the participation of the learners in the debriefing process, as well as their motivation to engage in self-reflection (Fanning & Gaba, 2007). Recent evidence suggests that self-debriefing using video-recordings of the simulated activity is equivalent to facilitator-led debriefing with regards to skill acquisition, as shown in two randomized controlled trials of OR crisis simulations to teach nontechnical skills (Boet et al., 2011; Boet et al., 2013). It should be noted, however, that in the studies by Boet et al. (2011 and 2013), trainees were provided with a list of essential nontechnical skills and learning points to guide the process of analytic reflection.

Regarding the design of simulation-based training, several aspects should be considered. Cook and coworkers (2012), in a meta-analysis comparing simulation-based interventions with regards to instructional methods used, identified repetitive practice and distributed practice as distinct design features that improved both knowledge and skill outcomes. Design features that were linked to enhanced skill outcome of simulations were cognitive interactivity, feedback, and individualization of the simulations (Cook et al., 2013).

The content of simulations should be aligned with learning objectives and needs, as well as the level of training of the participants (Beaubien & Baker, 2004). The simulated environment should be a realistic representation of the real-world work environment, in our case the OR, to allow trainees to immerse in the scenario and make meaningful experiences that can be generalized and applied to the real world. In design recommendations for simulation-based training in aviation, Rehmann (1995) has defined different facets of realism, or fidelity that influence the success of a simulation. Examples are equipment (e.g., functionality of equipment), environment (i.e., does the environment look like the “real thing”?), task (e.g., do tasks that are required of trainees in the simulation reflect real-life tasks in the respective workplace?), and psychological fidelity (i.e., does the trainee perceive the simulation as realistic?) (Rehmann, 1995). Psychological fidelity is critical for the
success of the simulation as it determines whether a trainee can fully immerse in the scenario and make meaningful experiences that induce learning (Beaubien & Baker, 2004). Multiple studies have assessed the fidelity of OR simulations based on trainees’ reactions, with positive responses from participants (Black et al., 2006; J. Paige et al., 2007; Powers et al., 2008; Undre, Koutantji, et al., 2007). Recurrent features of most studies involving OR team simulations include a realistic OR environment with a functional anaesthesia machine and an operating table, as well as a scrub table with instruments relevant for the simulated procedure; a simulated patient, usually a human patient manikin (e.g., SimMan, Laerdal Medical, Stavanger, Norway) with simulated vital functions (breath sounds, palpable pulses etc.) that can be modified from the control room; and OR team members, either represented by actual professionals from the respective specialty, or played by trained actors. Depending on the context, a surgical model allowing for completion of full or part tasks may also be part of the setup, ranging from virtual reality laparoscopy simulators (J. Paige et al., 2007) to realistic models of an anatomic region (Black et al., 2010). Some studies have reported successfully using trained actors in the roles of patients in simulations that required interaction with the patient, e.g. when simulating procedures under local or regional anaesthetic (Black et al., 2006; Kneebone et al., 2006).

**Debriefing and feedback.** The role of debriefing is well understood in the context of simulation-based training, as elaborated above. During everyday practice in the OR, however, the importance of debriefing and feedback appears to be less clear. As described above, debriefing in an educational context refers to a dialogue between trainer and trainee with the purpose of analyzing performance in view of target skills and behaviours; identifying gaps in performance; investigating possible causes for these gaps; and closing gaps by making a plan to modify behaviour in future encounters (Fanning & Gaba, 2007; Rudolph, Simon, Raemer, & Eppich, 2008). The role of the teacher in the debriefing dialogue is that of a facilitator to encourage the trainee to reflect on performance (Fanning & Gaba, 2007). Of note, debriefing for the purpose of the present section of this chapter refers to the educational activity as formative assessment involving only the trainer and the trainee, and is to be distinguished from the formal team debriefing that is conducted at the end of an operation to optimize team and organizational processes. Although different approaches to debriefing have been reported, common elements can be identified across authors.
Regarding the context of the debriefing, common recommendations include creating a nonthreatening environment and clarifying expectations regarding objectives and goals of the debriefing (Arora, Ahmed, et al., 2012; Fanning & Gaba, 2007; Rudolph et al., 2008). The debriefing itself is often subdivided in phases. In the initial reaction phase, the trainee is encouraged to describe what happened and how he or she perceived the experience (Arora, Ahmed, et al., 2012; Fanning & Gaba, 2007; Rudolph et al., 2008). Arora et al. (2012) pointed out the importance of engaging the learner to participate in the debriefing session, for instance by use of open-ended questions. In the analysis phase, the teacher helps the trainee to describe and diagnose relevant aspects of performance and to identify gaps and deficiencies as described above (Arora, Ahmed, et al., 2012; Rudolph et al., 2008). Ideally, the trainee engages in self-reflection with as little facilitation by the trainer as possible (Arora, Ahmed, et al., 2012; Fanning & Gaba, 2007). In the summary or diagnosis phase learning points are derived from identified gaps in performance, or positive aspects of performance are reinforced, in order to make a plan to enhance future performance (Arora, Ahmed, et al., 2012; Rudolph et al., 2008). Variable results have been reported regarding the use of additional instructional methods in combination with debriefing to enhance the learning outcome (Raemer et al., 2011). In a recent randomized controlled trial involving anaesthesiology residents in simulated OR crisis scenarios who were assigned to receive either no debriefing, oral debriefing, or oral debriefing enhanced with video-feedback, both oral and video-enhanced debriefing resulted in significant improvements in nontechnical performance, while simulation with no debriefing was found to be ineffective (Savoldelli et al., 2006).

Feedback has been described as “information that a system uses to make adjustments in reaching a goal” (p.777) (Ende, 1983). Feedback is different from debriefing as it is not so much a reflective process, but rather an exchange of information regarding performance. In the context of postgraduate education, feedback provides trainees with information about their performance, both negative and positive, the former to correct mistakes and address deficiencies, the latter to reinforce good performance (Ende, 1983). This information is important since recent evidence suggests that the ability of physicians to accurately assess own performance is often limited (Davis et al., 2006). This is particularly true for the self-assessment of nontechnical skills, which are often over- or underestimated by surgeons.
(Arora et al., 2011; Moorthy, Munz, Adams, Pandey, & Darzi, 2006). In the experiential learning process, reflection on inaccurate perceptions of own performance in the OR can potentially result in learning the wrong things.

Problems with feedback in postgraduate education have been known for several decades and led researchers to the definition of guidelines to improve feedback (Ende, 1983). Based on own observations and a review of the literature, Jack Ende (1983) recommended feedback to be informed by actual observations and focused only on observed performance. Furthermore, feedback should accurately describe observed performance and use pre-defined goals to objectively identify positive and negative behaviours without making generalized judgments about the trainee (Ende, 1983). Feedback should be kept at a manageable amount and address only those behaviours amenable to change by the trainee (Ende, 1983). Ende (1983) further emphasized the importance of the timing and setting for feedback, which should be agreed upon by both the teacher and the trainee, as unexpected feedback (especially when negative) may induce a negative affect in the trainee and impede learning (Ende, 1983).

In a recent study on the timing of feedback, the expectation of proximate feedback was associated with improved performance among a sample of 271 students, suggesting an advantage of timely feedback (Kettle & Haubl, 2010). In a recent observational study on debriefing and feedback as instructional methods in surgery, the authors identified these educational activities in only 46 percent of 35 observed cases (Ahmed, Sevdalis, Vincent, & Arora, 2013). Moreover, feedback was almost exclusively limited to the intraoperative phase and aimed at technical aspects of performance, while nontechnical skills were not addressed (Ahmed et al., 2013). In interviews with trainees and staff surgeons, Ahmed et al. (2013) identified perceived time constraints as main barrier to feedback and debriefing. Moreover, trainees frequently felt that feedback was perceived as a chore by staff surgeons, while the latter thought residents had no interest in receiving feedback (Ahmed et al., 2013). These misconceptions underscore Ende’s (1983) proposition that feedback should be agreed upon and planned by trainee and teacher together.
Despite representing distinct instructional methods, the terms “debriefing” and “feedback” are often used interchangeably in postgraduate education (Ahmed et al., 2013). In practice, the combination of the two methods may be beneficial, as feedback is often necessary to stimulate or encourage active reflection on performance, or to provide additional information to trainees during the analysis phase of debriefing (Arora, Ahmed, et al., 2012; Rudolph et al., 2008). Feedback on nontechnical aspects of performance in the OR was recently described as a feasible approach to the workplace-based teaching of nontechnical skills, and utility of this kind of teaching was perceived high by both educators and trainees (Spanager, Dieckmann, Beier-Holgersen, Rosenberg, & Oestergaard, 2015). Yule and coworkers (2008) conducted a trial to evaluate the feasibility of debriefing as an approach to teaching trainees about nontechnical aspects of operative performance. After receiving designated training in the observational assessment of nontechnical skills, participating surgeons assessed trainee performance in the OR using the NOTSS rating system, discussing nontechnical skills with trainees in individual postoperative debriefing sessions based on their observations of the trainee (Yule, Flin, Maran, Youngson, et al., 2008). Although participating surgeons felt positively about the utility of the debriefings based on NOTSS, the authors reported a relatively high attrition rate of 50 percent among staff surgeons which may reflect a lack of motivation to conduct debriefings on a regular basis in the OR (Yule, Flin, Maran, Youngson, et al., 2008). Despite positive reports regarding feasibility, however, data regarding the effectiveness of feedback or debriefing on nontechnical performance in the OR is lacking. With regards to technical skills, targeted feedback in the OR has been shown to result in superior improvements in operative performance when compared with practice without feedback (Grantcharov, Schulze, & Kristiansen, 2007).

In summary, debriefing and feedback represent two evidence-based instructional methods to enhance learning in the OR. By following best-practice recommendations, the effectiveness of debriefing/feedback can be maximized. The role and effectiveness of debriefing and feedback as instructional methods to teach nontechnical skills in the workplace is yet to be evaluated.
1.2.3.6 Evaluation of Training Interventions and Programs: The Kirkpatrick Framework

The evaluation of training interventions and programs serves several purposes. The primary goal of evaluation is to demonstrate effectiveness of the training in achieving predefined outcomes, thus to convince stakeholders that the investment of time, as well as human and financial resources is justifiable by the results (Kirkpatrick & Kirkpatrick, 2006). Training interventions in healthcare and postgraduate medical education targeted at nontechnical skills to improve efficiency and patient safety in the OR concern a large group of stakeholders, including trainees and trainers, who invest time and effort by attending or administering the training; hospital administrators who allocate designated time for training of health professionals; residency program directors who need to ensure curriculum time is used efficiently and residents attain all required competencies; and patients, who are entitled to high-quality and safe operative care. In addition to demonstrating the merit and value of training, evaluation also has the goal to identify problems and deficiencies in training design in order to modify and improve programs and interventions (Kirkpatrick & Kirkpatrick, 2006).

To date, Donald Kirkpatrick’s (1976) four-level framework remains the most commonly used approach to the evaluation of training in business, industry and academia (Alliger, Tannenbaum, Bennett, & Traver, 1997; Kraiger, Ford, & Salas, 1993; Salas, Wilson, Burke, & Wightman, 2006). Virtually all training interventions and programs reviewed in the present chapter thus far have used Kirkpatrick’s (1976) taxonomy to assess effectiveness of training. The four levels are: (1) reactions; (2) learning; (3) behaviour; (4) results (Kirkpatrick, 1976). On the first level, reactions, the satisfaction of training participants is assessed using methods such as post-course critiques and questionnaires (Kirkpatrick & Kirkpatrick, 2006). Reactions are at the basis of the assessment hierarchy, as the perceived usefulness does not guarantee learning and achievement of the overall training goal. However, Kirkpatrick and Kirkpatrick (2006) argue that trainees are less likely to accept new concepts and thus learn if the training is perceived as useless, or a waste of time (Kirkpatrick & Kirkpatrick, 2006). The evaluation of participants’ reactions to the training may also yield valuable information about potential problems and areas for improvement,
which is particularly relevant if the program fails to deliver the desired outcomes 
(Kirkpatrick & Kirkpatrick, 2006). Alliger et al. (1997) have subdivided Kirkpatrick’s 
(1976) reactions level into three sub-levels in order to allow for a more granular evaluation. 
The authors differentiated between reactions concerning affect, i.e., how participants liked 
or enjoyed the training; reactions pertaining to the perceived utility of the training; and 
combined reactions (Alliger et al., 1997).

Participant learning is evaluated at the next level of Kirkpatrick’s (1976) framework. 
Traditionally, learning was assessed as knowledge pertinent to the area of training 
(Kirkpatrick, 1976). Kraiger et al. (1993) have proposed a more in-depth characterization of 
learning evaluation, comprising cognitive outcomes, skill-based outcomes and affective 
outcomes of learning. Cognitive outcome includes declarative knowledge and procedural 
knowledge, and is commonly assessed through multiple-choice questions and oral or written 
tests (Kraiger et al., 1993). Skill-based outcomes pertain to technical aspects of 
performance, such as automaticity (Kraiger et al., 1993). Affective outcomes comprise 
attitudinal changes and motivational issues, e.g., the readiness and willingness of trainees to 
change behaviour (Kraiger et al., 1993). A shift to more positive attitudes about the learning 
objectives (e.g., nontechnical skills, patient safety) as a result of a training intervention may 
thus indicate the trainees’ motivation to implement the newly acquired concepts into daily 
practice. The evaluation of learning ensures that trainees have acquired the necessary 
knowledge and are willing to change relevant behaviour (Kirkpatrick & Kirkpatrick, 2006).

Behaviour is assessed at the third level of Kirkpatrick’s (1976) hierarchy. Alliger et al. 
(1997) defined Kirkpatrick’s (1976) behaviour level as transfer of training and underscore 
the impact on performance in the workplace (e.g., the OR). In addition to the willingness to 
change and the required knowledge, Kirkpatrick and Kirkpatrick (2006) emphasize the role 
of organizational factors in the success of a training intervention. In particular, the climate 
within an organization may impede or enhance behaviour change and transfer of training. 
For instance, when a participant in a nontechnical skills course is not encouraged by his or 
her supervisor to apply learned behaviours in the OR, he or she is less likely to change 
behaviour as a result of the training. Conversely, if new behaviours are encouraged by role
models, and leaders within the organization welcome and support the change, the training is more likely to result in behavioural change (Kirkpatrick & Kirkpatrick, 2006).

The highest level of evaluation in Kirkpatrick’s (1976) framework relates to results, or outcome at the level of the organization. In the context of nontechnical skills training, results may encompass enhanced efficiency in perioperative processes such as a decrease in delays or disruptions of operative flow (e.g., through a reduction of communication failures, or improved coordination of team activities in the OR). The most important and desirable results, however, would be in patient safety, measureable as a decrease in perioperative complications or death.

1.2.3.7 Assessment of Nontechnical Skills

Evaluation, both in the context of program or training effectiveness, and for the purpose of formative feedback or promotion and certification in postgraduate education, relies on the assessment of trainees’ relevant knowledge, skills and attitudes. Norcini et al. (2011) proposed several criteria for good assessment in postgraduate education, including the reproducibility or consistency of the assessments; feasibility in a particular context or environment; educational effect and benefit of the assessment; acceptability of assessments to stakeholders including the healthcare system, regulatory bodies, educators and residency programs, patients, and the trainees themselves; and validity of the results of an assessment. Validity refers to the degree to which test scores generated in an assessment represent a trainee’s competence or performance in the construct that is being evaluated.

1.2.3.8 Validity

The unified concept of validity, as proposed by Samuel Messick (1989) and endorsed in the 1999 Standards for Educational and Psychological Testing, combines and interrelates several aspects of validity to a comprehensive understanding of construct validity (Messick, 1995). Samuel Messick (1989) proposed this definition of validity:
“Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment” (p.13) (Messick, 1989).

Following the 1999 Standards for Educational and Psychological Testing which are based on Messick’s (1995) unitary framework, five sources of validity evidence can be explored to support construct validity of test scores: evidence based on content; response process; internal structure; relations to other variables; and consequences of testing (AERA, APA, & NCME, 1999). It is important to understand that validity in the sense of the unitary framework does not relate to the test instrument, but to the scores and their interpretation in the context of the assessment (Cook & Beckman, 2006)

**Evidence based on content:** This source of validity evidence reflects the adequacy of the test content (e.g., test items, questions, or tasks) in representing the underlying construct that is being tested (AERA, APA, & NCME, 1999). The development of test items by individuals who are content experts in the domain of the test is an example of validity evidence based on content (Cook & Beckman, 2006). Construct definition in a test or instrument is a further aspect of content evidence (Cook & Beckman, 2006).

**Evidence based on response process:** Response process relates to factors associated with taking or administering the test, or conducting assessments (Downing, 2003). Evidence in this category can be gathered from comments and responses of individuals taking a test (examinees), or administering or using a test (e.g., examiners, observers) about their interpretation and understanding of test items (e.g., to ensure examiners are applying the appropriate criteria and do not make judgments based on wrong criteria) (AERA, APA, & NCME, 1999). Further examples of response process evidence are the accuracy of descriptions regarding the use of the test and the rationale for the scoring method used in an assessment (Downing, 2003). Response process evidence is also generated when test items are revised to improve clarity (e.g., wording, definitions) based on comments from test-taker or test-users. Rater training (e.g., training of observers prior to the assessment of
nontechnical performance in the OR) represents another important aspect of response process evidence, as poorly trained raters are less likely to generate valid scores (Cook & Beckman, 2006). Consequently, the appropriate use of test items and criteria by examiners when assessing a trainee is another example for validity evidence based on response process (AERA, APA, & NCME, 1999).

**Evidence based on internal structure:** Internal structure evidence describes the psychometric properties of test questions, assessment tools or rating scales. Psychometric properties contributing to internal structure evidence include reliability, internal consistency, and generalizability. Reliability is a measure of reproducibility and dependability of test scores (Downing, 2004). One aspect of reliability is the stability of test scores over time, measured as test-retest reliability. A more important measure of reliability is the agreement between different examiners when scoring the same examinee (Cook & Beckman, 2006). Agreement between examiners or raters should be high to ensure scores reliably reflect the performance of the examinee in the construct being assessed. Agreement between raters can be measured as a percentage of identical scores, or as a correlation between raters’ scores (Cook & Beckman, 2006). The most commonly used measure of inter-rater agreement though is the intra-class correlation coefficient (ICC). Although reliability alone does not guarantee validity of test scores, it is a fundamental component of the validation process, since test scores that are neither reproducible nor dependable do not allow for meaningful interpretation of performance (Crossley, Humphris, & Jolly, 2002).

The generalizability coefficient is a further measure of reliability. The generalizability coefficient represents an estimate of the degree to that test scores represent true variance in examinees’ performance related to the construct being tested, and are not the result of rater error or other factors (e.g., stringency or leniency of different raters, differences in difficulty between test scenarios) (Crossley, Davies, Humphris, & Jolly, 2002).

Internal consistency, or scale reliability, of a rating tool describes the degree to that different items of the tool measure the same underlying construct. Internal consistency of a rating tool can be assessed by computing correlations between different items of the tool in inter-item correlations. Correspondingly, item-total correlations measure the relationship between
individual items and the total score, i.e., the overall construct being measured (Downing, 2003). A commonly used statistical method to estimate internal consistency of a test or rating scale is the Cronbach’s alpha coefficient (Downing, 2004). The Cronbach’s alpha coefficient reflects the correlation between scores from different assessments using the same test or tool.

**Evidence based on relations to other variables:** This source of validity evidence relates to the (positive or negative) correlation of test scores with scores from other tests that represent the construct that is being tested, or a different construct (AERA, APA, & NCME, 1999). An example of a convergent correlation is the correlation of test scores generated with a new assessment tool with scores obtained from an older, well-established tool measuring the same construct (Downing, 2003). Divergent correlations between test scores and scores from measurements of a different construct equally contribute to the validity evidence. Further examples of validity evidence in this category are the relations of test scores with specific criteria of performance, and the generalizability of test score-criterion relationships across different settings. The latter source of validity evidence can be generated by comparing different validation studies with regards to the score-criterion relationship (AERA, APA, & NCME, 1999).

**Evidence based on consequences of testing:** This category describes potential effects of the assessment on examinees or examiners. More specifically, evidence is gathered in this category to ensure the testing or assessment itself has no negative effects (Downing, 2003). An example for consequences of testing is an ultra-high-stakes exam that may have a major impact on an individual’s career and life (e.g., failing the board certifying exams after investing 6 years in residency training). On the other hand, passing a candidate with insufficient knowledge or clinical skills may result in harm to future patients (Downing, 2003). Thus, a possible source of validity evidence regarding consequences of testing in high-stakes assessment may be the definition of sound, evidence-based pass-fail criteria (Downing, 2003). A problem with this category of validity evidence, however, is its relative subjectivity, since pass-fail criteria are usually based on expert opinions and less frequently on evidence.
In summary, the validation of test scores obtained from a specific test or rating tool is a continuous process with the goal of obtaining dependable and reproducible scores that represent an individual’s performance in the construct that is being tested. A central argument in the unified concept of validity is that validity solely describes properties of test scores, i.e., how well one can rely on and trust the scores to provide an adequate representation of the examinee’s performance in the construct tested. Validity does not describe properties of the assessment tool. Thus, an assessment tool cannot be valid in itself, but validity evidence must be collected in the exact context the tool is used in order to ensure validity of test scores (Cook & Beckman, 2006). Validation, however, does not necessitate the assessment of all possible sources of validity evidence every single time a test is administered. Rather, validity evidence from previous studies using the test or tool in the proposed context can also inform the validity argument. In the following section, I will introduce the most pertinent tools for the assessment of surgeons’ nontechnical skills in the OR and review the validity evidence in the literature that supports the adequacy of test scores obtained with each tool.

1.2.3.9 Assessment Tools of Nontechnical Skills in the OR

Assessment in postgraduate education has the purpose of documenting the adequate development of knowledge and skills to make decisions on promotion during residency training, and to document the attainment of required competencies at the end of residency. The latter summative assessment forms the basis for decisions regarding the successful graduation from residency training and the attainment of board certification. The question arises how to obtain a fair and valid assessment of surgical residents’ nontechnical competencies, specifically with regards to nontechnical skills in the OR.

The assessment of nontechnical or CRM skills was pioneered in aviation and traditionally comprised the observation and rating of flight crews’ behaviours in the cockpit (Helmreich et al., 1999). Frameworks of skills pertaining to the construct of CRM or nontechnical skills formed the basis for the assessments, while behavioural markers, i.e., examples for positive and negative behaviours pertaining to a skill item, facilitated the rating (Helmreich &
Foushee, 2010). This methodology of assessment based on observable behaviours was emulated in the development of rating systems for nontechnical skills in the OR. Early and ground breaking work on behavioural markers was reported by Helmreich and Schaefer (1994) with their checklist of CRM skills and behaviours observed in the OR, as well as de Leval et al.’s (2000) and Carthey et al.’s (2003) behavioural markers of surgical excellence. The checklists and frameworks of skills and behaviours reported in those studies, however, were not used in subsequent studies for the assessment of performance of surgical teams in the OR.

The vast majority of frameworks and rating tools of nontechnical skills used in medicine today are based on work in aviation human factors and follow a basic framework comprising a hierarchy of categories, elements, and observable behaviours (Klampfer et al., 2001) At the top level of the hierarchy are the categories, representing different aspects or areas of skills that are relevant for the overarching construct that is being assessed. Each category is subdivided into a number of skill elements that conceptually belong to the respective category but represent distinct aspects of performance within the category (Flin et al., 2003). Finally, elements are described by a number of behavioural markers, which represent observable behaviours (positive or negative) of workers that enhance (or impede) relevant aspects of performance in the corresponding work environment (Klampfer et al., 2001).

A schematic of this generic framework is shown in Figure 6. The definition of performance in a particular framework depends on the conceptualization of the underlying construct, such as safety and efficiency in the work environment in the construct of nontechnical skills (Flin et al., 2003). Behavioural markers are based on empirical research in the respective work environment and can be derived from retrospective data analysis of incidents and accidents, structured observations in the workplace and during simulations, task analysis, surveys, and focus group interviews (Klampfer et al., 2001). A general prerequisite of skill categories, elements, and behavioural markers pertaining to the elements is that they should be as mutually exclusive as possible, while still representing the overarching construct (Flin et al., 2003; Klampfer et al., 2001). Moreover, behaviours pertaining to elements and categories should be observable in the work environment, or inferable from communication between workers (Klampfer et al., 2001).
Of note, the number of categories, elements and behavioural markers varies depending on the conceptualization of the underlying construct and the design of the individual rating system. Moreover, behavioural markers may comprise both positive and negative examples, or be limited to positive examples of behaviour.

Figure 6: Basic framework of rating systems for the assessment of performance based on observable behaviours in high-risk work environments.


In the following section, I will introduce the most commonly used and reported rating systems and taxonomies of non-technical skills in the OR environment, and analyze each
rating system with regards to evidence supporting the validity of scores generated with the system. Of note, I will limit this review to rating systems aimed at, or including, surgical teams in the OR environment, as this is the focus of the present work.

1.2.3.9.1 Non-Technical Skills for Surgeons (NOTSS)

NOTSS represents a taxonomy of nontechnical skills and behavioural markers as well as a rating framework for nontechnical skills of surgeons in the OR. In contrast to the majority of rating systems that are focused on entire OR teams, NOTSS was specifically designed to assess individual surgeons’ nontechnical performance in the OR environment (Yule, Flin, Paterson-Brown, Maran, et al., 2006). The design process described by Yule et al. (2006) involved a literature review of nontechnical skills relevant for surgeons in the OR (Yule, Flin, Paterson-Brown, & Maran, 2006), a cognitive task analysis, surveys of OR personnel regarding attitudes about teamwork and safety (Flin, Yule, McKenzie, Paterson-Brown, & Maran, 2006), and critical incident interviews with surgeons (Yule, Flin, Paterson-Brown, Maran, et al., 2006). In the interviews, surgeons were asked about challenging cases with a focus on surgeons’ positive and negative behaviours associated with the events (Yule, Flin, Paterson-Brown, Maran, et al., 2006). Human factors experts extracted 150 behaviours from the interviews which were subsequently grouped in a skill taxonomy (Yule, Flin, Paterson-Brown, Maran, et al., 2006). In an iterative process involving surgeons and psychologists the skills were organized in a framework comprising 5 categories of skills (situation awareness; decision making; task management; leadership; and communication and teamwork) and 14 skill elements. Last, behavioural markers were defined as observable positive and negative behaviour to guide rating of skill at the level of element and categories (Yule, Flin, Paterson-Brown, Maran, et al., 2006). The rating scale was designed as a four-point categorical scale with the markers “4 good”, “3 acceptable”, “2 marginal”, and “1 poor”. A “not observed” score was added to account for situations where a particular element or category was not required and could for that reason not be observed (Yule, Flin, Paterson-Brown, Maran, et al., 2006). When Messick’s (1995) framework is applied to analyze the development process of NOTSS involving literature reviews, surveys, task analysis, and cognitive interviews with experts, substantial evidence can be identified in the content category and, to a lesser degree,
the category response process to support construct validity of scores generated with the rating system (AERA, APA, & NCME, 1999). Psychometric properties of the rating system were subsequently assessed in a study of 44 surgeons from different specialties who used the framework to rate performance of surgeons in 11 scripted OR videos (Yule, Flin, Maran, Rowley, et al., 2008). Inter-rater agreement was assessed by calculating mean within-group agreement, as well as ICCs (Yule, Flin, Maran, Rowley, et al., 2008). Mean within-group agreement was found to be between 0.51 and 0.72, which was interpreted by the authors as acceptable, although not meeting the predefined minimum of > 0.7 for an acceptable level of agreement (Yule, Flin, Maran, Rowley, et al., 2008). ICCs for average measures of all raters were excellent (0.95 - 0.99), while ICCs calculated for single raters were comparatively low (0.29 - 0.66) (Yule, Flin, Maran, Rowley, et al., 2008). Agreement of raters with reference ratings from human factors experts was > 60 percent for all skill categories (Yule, Flin, Maran, Rowley, et al., 2008). Some effects of rater specialty and context of the OR video were described to influence within-group agreement. Lastly, Yule et al. (2008) found high internal consistency between category and element scores, indicating alignment of categories and elements with the overall construct of nontechnical skill. Following feedback from the raters, the category “task management” was removed from the rating framework resulting in the final version of NOTSS (Table 2) (Yule, Flin, Maran, Rowley, et al., 2008). With regards to construct validity, response process evidence was added through designated rater training and the use of rater feedback to modify the tool (Yule, Flin, Maran, Rowley, et al., 2008). Internal structure evidence can be derived from the analysis of the psychometric properties (Table 3). Differences between ratings from novice and expert raters were shown in a different study, with novices scoring surgeons’ nontechnical performance in scripted videos lower than experts (Yule et al., 2009). Based on the results, the authors highlighted the importance of rater training and calibration to achieve reliable scores with NOTSS (Yule et al., 2009).

The authors subsequently tested NOTSS in the real OR environment in a study assessing the feasibility of nontechnical skill debriefing (Yule, Flin, Maran, Youngson, et al., 2008). Study participants were staff surgeons who were trained in the use of NOTSS in three-hour group sessions comprising audio-visual didactic material and scripted videos to practice rating (Yule, Flin, Maran, Youngson, et al., 2008). Surgeons then applied NOTSS to assess
and rate performance of their trainees during cases in the OR, and used the NOTSS ratings in individual postoperative debriefing sessions to provide feedback to trainees regarding nontechnical performance (Yule, Flin, Maran, Youngson, et al., 2008). Of 22 surgeons who were initially recruited, only 11 completed the study. Surgeons used NOTSS to observe and rate trainee performance in a total of 43 cases that were performed by the trainee (Yule, Flin, Maran, Youngson, et al., 2008). In an anonymous questionnaire, participating surgeons indicated that they found NOTSS useful as tool for debriefing and feedback (Yule, Flin, Maran, Youngson, et al., 2008). Perceived difficulties in the use of NOTSS were thought to be owing to a lack of training in the use of the tool and the rating process itself. Surgeons also found it difficult to rate while being scrubbed themselves, although these findings were not quantified (Yule, Flin, Maran, Youngson, et al., 2008). A limitation of this study was the low participation rate (50 percent) resulting in a potential bias through the selection of surgeons who were particularly supportive of the concept of nontechnical skills. Validity evidence from this study contributed to the category response process (rater training and feedback from raters), as well as consequences (use of the scores to teach trainees).
<table>
<thead>
<tr>
<th>Categories</th>
<th>Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gathering information</td>
</tr>
<tr>
<td></td>
<td>Understanding information</td>
</tr>
<tr>
<td>Situation awareness</td>
<td>Projecting and anticipating future state</td>
</tr>
<tr>
<td>Decision making</td>
<td>Considering options</td>
</tr>
<tr>
<td></td>
<td>Selecting and communicating options</td>
</tr>
<tr>
<td></td>
<td>Implementing and reviewing decisions</td>
</tr>
<tr>
<td>Communication and teamwork</td>
<td>Exchanging information</td>
</tr>
<tr>
<td></td>
<td>Establishing a shared understanding</td>
</tr>
<tr>
<td></td>
<td>Coordinating team activities</td>
</tr>
<tr>
<td>Leadership</td>
<td>Setting/maintaining standards</td>
</tr>
<tr>
<td></td>
<td>Supporting others</td>
</tr>
<tr>
<td></td>
<td>Coping with pressure</td>
</tr>
</tbody>
</table>

**Table 2: Non-Technical Skills for Surgeons (NOTSS) rating framework, version 1.2.**


Crossley and coworkers (2011) used NOTSS in an observational study to assess nontechnical performance of 85 surgical trainees in the OR. In 404 operations across six surgical specialties, 100 trained raters observed the trainees during operations and rated nontechnical performance on the NOTSS system, resulting in 715 assessments (Beard, Marriott, Purdie, & Crossley, 2011). The raters, the majority of whom were
anaesthesiologists (n=56) and scrub nurses (n=39), received brief training in the use of NOTSS (Crossley et al., 2011). In end-of-study questionnaires, the majority (75 percent) of participating raters found NOTSS useful to assess nontechnical performance (Crossley et al., 2011). Within the NOTSS, categories of interpersonal skills (communication and teamwork; leadership) were perceived to be easier to rate than cognitive skills (situation awareness; decision making) (Crossley et al., 2011). Correlations with other variables showed significant correlations of all four NOTSS domain scores (domain score = the sum of a category score and its element scores) with PBAs that were concurrently conducted by staff surgeons. Moreover, NOTSS domain scores were correlated with scores on the Objective Structured Assessment of Technical skills (OSATS) scores of trainees, suggesting a link between technical and nontechnical performance (Crossley et al., 2011). Of note, decision making on NOTSS was most strongly correlated with technical skill as measured on OSATS (Crossley et al., 2011). Nontechnical performance was also found to be correlated with years of training in the U.K., but not with the surgeons’ age or training outside of the U.K. (Crossley et al., 2011). This was an interesting finding, which may indicate an influence of the efforts in U.K. training programs to include formal assessments of nontechnical skills in the ISCP (ISCP, 2013b).

Reliability was assessed in a generalizability study (G-study) through the analysis of different sources of variance (Crossley et al., 2011). Crossley et al. (2011) found nontechnical performance of the trainees (i.e., the “true” variance) to contribute 31 percent to the total variance, while rater factors such as stringency and leniency, as well as subjectivity (i.e., errors), accounted for 27 percent and 20 percent, respectively (Crossley et al., 2011). Results of the G-study were used for reliability modeling in a D-study to determine the number of cases for a given number of assessors that resulted in an acceptable reliability coefficient (a coefficient of > 0.8 was considered acceptable by the authors) (Crossley et al., 2011).

Of note, nine percent of scores were in the “not applicable” category of the scale. This was most commonly found to be the case in the category of leadership. Field notes provided an explanation for this finding, with raters finding it difficult to score leadership of trainees when staff surgeons took over the cases (Crossley et al., 2011). In contrast to the study by
Yule et al. (2009), expert ratings were found to be more stringent than novice scores (Crossley et al., 2011). In summary, Crossley et al. (2011) generated response process evidence (feedback from raters on feasibility), internal structure evidence (generalizability study and D-study), and evidence pertaining to the relationship to external variables (correlations) (Table 1). Two studies were identified that applied NOTSS to rate surgeons’ performance in a simulated OR environment involving intraoperative crisis situations (Arora et al., 2011; Lee et al., 2012). In both studies, trainees were found to misjudge their performance when compared with expert ratings (Arora et al., 2011; Lee et al., 2012). Finally, NOTSS was recently translated and modified to assess nontechnical skills of Danish surgeons (Spanager et al., 2012). Good reliability of the NOTSSdk was documented by the group, both for assessments of recorded OR simulations (Spanager et al., 2013), and for observation in the real OR (Spanager, Konge, et al., 2015).

<table>
<thead>
<tr>
<th>Reference</th>
<th>Context of assessment</th>
<th>Validity evidence following the unitary framework of construct validity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Content</td>
</tr>
<tr>
<td>Yule et al.</td>
<td>n/a</td>
<td>+</td>
</tr>
<tr>
<td>(2006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yule et al.</td>
<td>Scripted videos</td>
<td>-</td>
</tr>
<tr>
<td>(2008a)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yule et al.</td>
<td>Scripted videos</td>
<td>-</td>
</tr>
<tr>
<td>(2008b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yule et al.</td>
<td>OR</td>
<td>-</td>
</tr>
<tr>
<td>(2009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crossley et al.</td>
<td>OR</td>
<td>-</td>
</tr>
<tr>
<td>(2011)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee et al.</td>
<td>Simulation</td>
<td>-</td>
</tr>
<tr>
<td>(2012)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Validity evidence supporting construct validity of scores generated with the NOTSS system.
The combined validity evidence from the NOTSS studies reviewed here supports construct validity of scores obtained in simulation as well as the real OR environment. Findings suggest that valid test scores can be obtained with NOTSS when the tool is used by well-trained raters, both in simulated scenarios and in the real OR. Difficulties in the use of NOTSS were identified when rating cognitive skills (both in simulation and the real OR). In the real OR environment, the presence of staff surgeons may impede the rating of trainee performance with a potential marginalization of the trainee role due to staff dominance and takeover.

1.2.3.9.2 Oxford NOTECHS

The Oxford NOTECHS is a rating system of nontechnical skills that was specifically designed to assess OR teams in the workplace (Mishra, Catchpole, & McCulloch, 2009). The design of the Oxford NOTECHS was based on the aviation NOTECHS and closely resembles its framework of four categories (leadership and management; teamwork and cooperation; problem-solving and decision making; and situation awareness) (Flin et al., 2003). Each category is complemented with elements of nontechnical skills (between 3 and 5) (Mishra et al., 2009). Between 3 and 7 behavioural markers provide examples of (only positive) behaviour pertaining to the elements (Mishra et al., 2009). Categories and elements of the Oxford NOTECHS are shown in Table 4. Behavioural markers were defined based on the work by Carthey et al. (2003), and reviewed for relevance and adequacy by specialists in surgery and anaesthesiology, a human factors expert, and aviation CRM trainers (response process) (Mishra et al., 2009). In analogy to the aviation NOTECHS, performance is rated on the Oxford NOTECHS based on observable behaviour, using a four-point Likert-type scale with the scores “1 below standard”, “2 basic standard”, “3 standard”, and “4 excellent” (Mishra et al., 2009). The four rating options are explained by a generic anchor, describing the consequence of the behaviour for patient safety and teamwork (e.g., the anchor for “1 below standard” is: “behaviour directly compromises patient safety and effective teamwork”) (Mishra et al., 2009). In contrast to NOTSS and the aviation NOTECHS,
behaviour is only rated at the level of the categories, while elements are not scored (Mishra et al., 2009). Moreover, each sub-team in the OR (i.e., surgical team; nursing team; and anaesthesiology team) is rated separately. Additional behavioural markers specific to the team are provided as “sub-team modifiers” (Mishra et al., 2009).

<table>
<thead>
<tr>
<th>Category</th>
<th>Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leadership and management</td>
<td>Leadership, Maintenance of standards, Planning and preparation, Workload management, Authority and assertiveness</td>
</tr>
<tr>
<td>Teamwork and cooperation</td>
<td>Team building/ maintaining, Support of others, Understanding team needs, Conflict solving</td>
</tr>
<tr>
<td>Problem-solving and decision making</td>
<td>Definition and diagnosis, Option generation, Risk assessment, Outcome review</td>
</tr>
<tr>
<td>Situation awareness</td>
<td>Notice, Understand, Think ahead</td>
</tr>
</tbody>
</table>

**Table 4: Categories and elements of the Oxford NOTECHS system.**
Mishra et al (2009) applied the tool to observe teams in the OR, both before and after a training intervention targeted at nontechnical and CRM skills. Two observers (a surgeon and a human factors expert) were trained in the use of the tool and observed 65 laparoscopic cholecystectomies on the OR. Agreement between the two raters, computed as within-group agreement, was good (between 0.83 and 0.99) for all categories in three sub-teams (except situation awareness of the anaesthesiology team) (Mishra et al., 2009). Total scores were negatively correlated with technical errors that were recorded concurrently by a different observer, with a particularly strong negative correlation between NOTECHS scores of the surgical sub-team and error counts (Mishra et al., 2009). Moreover, team NOTECHS scores were significantly higher in the 39 observations that followed a CRM-based training intervention when compared with scores before the intervention (n=26) (Mishra et al., 2009). An important limitation of this finding, however, is the lack of blinding of the observers, who were involved in the training intervention (Mishra et al., 2009). To reduce the risk of bias, the authors introduced a third rater who was blinded to the training condition and rated 11 cases (Mishra et al., 2009). Good agreement between the unblinded raters and the third, blinded rater was interpreted by the authors to support the adequacy of the ratings (Mishra et al., 2009). The relationship to other variables was assessed by the concurrent use of the OTAS teamwork rating scale (Undre, Healey, Darzi, & Vincent, 2006). Oxford NOTECHS ratings from one observer correlated significantly and positively with OTAS scores generated by a different observer, who used OTAS concurrently with NOTECHS for a small number of cases (Mishra et al., 2009). The tool was subsequently used to assess nontechnical performance of OR teams before and after a CRM-based training intervention, showing good within-group agreement between raters (0.86 - 0.98) (McCulloch et al., 2009). Of note, these results included data from the previously discussed validation study (Mishra et al., 2009). The inverse correlation between nontechnical skills and technical errors, as demonstrated in the abovementioned study by Mishra et al (2009) was replicated in this study, with a particularly strong inverse correlation between surgical sub-team scores on the category situation awareness and technical error (McCulloch et al., 2009). As in the previous
study, observers were not blinded to the training condition, but a percentage of ratings were again conducted by a blinded observer, who was in good agreement with the two primary raters (McCulloch et al., 2009). Catchpole et al. (2010) used the Oxford NOTECHS to rate OR teams from maxillofacial, vascular and neurosurgery at three different hospital sites before and after a CRM-based training intervention. Two human factors experts observed a total of 112 operations. Again, agreement between raters, computed as within-group agreement, was good (> 0.7) for all categories of the scale (K. R. Catchpole et al., 2010). Comparison of scores before and after the training intervention did not reveal a significant effect of the training, but significant differences in scores between hospital sites were identified (K. R. Catchpole et al., 2010).

In summary, good validity evidence has been documented for the Oxford NOTECHS in the categories content, internal structure, and relationship with other variables, although the latter should be treated with caution due to the risk of observer bias (Table 5).

<table>
<thead>
<tr>
<th>Reference</th>
<th>Validity evidence following the unitary framework of construct validity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Content</td>
</tr>
<tr>
<td>Mishra et al. (2009)</td>
<td>+</td>
</tr>
<tr>
<td>McCulloch (2009)</td>
<td>-</td>
</tr>
<tr>
<td>Catchpole et al. (2010)</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5: Validity evidence supporting scores generated with the Oxford NOTECHS rating system.

1.2.3.9.3 Observational Teamwork Assessment for Surgery (OTAS)

The OTAS tool was designed specifically to measure team performance in the OR by observation (Healey, 2004). OTAS is based on a conceptualization of team performance as an input-process-output model, with process reflecting interactions among members of the
OR team, as well as interactions of team members with technology and patients (Healey, 2004). Various aspects of team behaviours influence the team process. Healy et al. (2004) based their framework of team behaviours on a published teamwork model comprising: team orientation; team leadership; communication; team monitoring; team feedback; backup behaviour; coordination (Dickinson & McIntyre, 1997). The OTAS framework for the surgical team during the intraoperative phase is shown in Table 6.

<table>
<thead>
<tr>
<th>Skill category</th>
<th>Exemplar behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Communication</strong></td>
<td>- Asks team if all prepared to begin the operation</td>
</tr>
<tr>
<td>Quality and quantity of</td>
<td>- Requests and instructions to team communicated clearly and effectively</td>
</tr>
<tr>
<td>information exchanged among</td>
<td>- Provides information to whole team on progress</td>
</tr>
<tr>
<td>team members</td>
<td>- Informs team of technical difficulties/changes of plan</td>
</tr>
<tr>
<td><strong>Coordination</strong></td>
<td>- Gives prior notification of requirements to team to enhance timing of tasks</td>
</tr>
<tr>
<td>Management and timing of</td>
<td>- Coordinate use of equipment, such as camera in minimal access</td>
</tr>
<tr>
<td>activities and tasks</td>
<td>- Contribute to smooth exchange of instruments and provisions with scrub nurse</td>
</tr>
<tr>
<td>**Cooperation/ back up</td>
<td>- Responds to requests and questions from nursing team</td>
</tr>
<tr>
<td>behaviour**</td>
<td>- Responds to requests or questions from anesthesia team</td>
</tr>
<tr>
<td>Assistance provided among</td>
<td>- Helps with smooth instrument exchange with scrub nurse</td>
</tr>
<tr>
<td>members of the team, supporting</td>
<td>- Supports Surgical group assistants and compensates for lack of experience</td>
</tr>
<tr>
<td>others, and correcting errors</td>
<td></td>
</tr>
<tr>
<td><strong>Leadership</strong></td>
<td>- Instructions and explanations provided to assistants</td>
</tr>
<tr>
<td>Management and timing of</td>
<td>- Advises anesthesia team or nursing team to call for additional help if required</td>
</tr>
<tr>
<td>activities and tasks</td>
<td>- Supervision provided for staff lacking familiarity with tasks or equipment</td>
</tr>
<tr>
<td></td>
<td>- Assertive in controlling noise and distractions in theatre</td>
</tr>
<tr>
<td>**Monitoring/ situational</td>
<td>- Asks anesthesia about patient condition</td>
</tr>
<tr>
<td>awareness**</td>
<td>- Asks scrub nurse if swabs, needles, and instrument count correct</td>
</tr>
<tr>
<td>Team observation and awareness</td>
<td></td>
</tr>
<tr>
<td>of ongoing processes</td>
<td></td>
</tr>
</tbody>
</table>
Although some behavioural markers of OTAS could potentially be observed in individuals, assessments using this tool are entirely focused on team performance. OTAS offers a brief definition of each behavioural construct, as well as positive exemplar behaviours pertaining to that behaviour (Healey, Undre, & Vincent, 2004). Thus, OTAS differs from NOTSS and NOTECHS not only in the assessment focus (teamwork rather than nontechnical skills), but also in its basic rating framework. Descriptive anchors that are provided to facilitate the rating on a seven-point Likert-type scale are defined in terms of the impact of the observed behaviour on team function (Healey, 2004). Of note, team function is assessed with OTAS for the preoperative, intraoperative and postoperative phase in the OR, with a version for the surgical team, nursing team and anaesthesiology team (with different behavioural markers for each sub-team and each time-point) (Healey et al., 2004). The initial version of OTAS comprised five behavioural constructs: communication; cooperation; co-ordination; shared leadership; and monitoring (Undre et al., 2006). In a feasibility study, the authors observed 50 general surgery operations, rating team behaviours on the rating scale and task completion on a checklist (Undre et al., 2006). In a subsequent observational study, feasibility was tested in 50 urology procedures, confirming the applicability of OTAS across different specialties (Undre, Sevdalis, Healey, Darzi, & Vincent, 2007). In a further observational study teams of raters observed 12 cases in the OR, six of which were observed by an expert-novice rater team, and six by two expert raters (Sevdalis et al., 2009). Analysis of scores revealed good correlation between expert-expert ratings, while expert-novice rater teams correlated poorly (Sevdalis et al., 2009). Hull et al. (2011) refined the tool (removal and modification of behavioural markers) based on an expert consensus, and conducted observations during 30 procedures using 2 trained raters. Raters reported high observability
of the new/modified markers, while agreement between the raters was found to be high (Cohen’s kappa > 0.41) for 109 of the 130 behavioural markers (Hull, Arora, Kassab, Kneebone, & Sevdalis, 2011). Inter-rater agreement, calculated as ICCs, was good (>0.7) for all behavioural constructs, except for the constructs co-ordination (ICC = 0.67) and team monitoring (ICC = 0.64) which were acceptable (Hull et al., 2011). Russ et al. (2012) described a training protocol for OTAS comprising an initial instruction in the tool, followed by the independent review and rating of three simulated procedures, and subsequent ratings in the real life OR. Four participating novice assessors and one expert rater independently rated 10 operations in the OR, blinded to each other’s ratings (Russ et al., 2012). After each operation, raters compared their notes and discussed ratings in a debriefing session (Russ et al., 2012). Analysis of scores form all ten cases showed that agreement between novice and expert ratings (as measured by ICC) improved significantly from the beginning to the middle stage of the training, as well as from the middle to the end of the study period, with all ICCs ranging around or over the 0.7 mark at the end of the observations (Russ et al., 2012). More recently, OTAS has been adapted for use in German and Latin American ORs. Acceptable reliability was documented for the German OTAS-D (all ICCs > 0.72) (Passauer-Baierl et al., 2014) and the Colombian OTAS-S (Amaya Arias et al., 2014). In summary, substantial evidence supporting construct validity of OTAS scores can be derived from the response process category, while internal structure evidence is reasonable (Table 7). Relationships to other variables and consequences were not explored in the studies identified for this review.

<table>
<thead>
<tr>
<th>Observational Teamwork Assessment for Surgery (OTAS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Healy et al. (2004)</td>
</tr>
<tr>
<td>+</td>
</tr>
<tr>
<td>-</td>
</tr>
<tr>
<td>-</td>
</tr>
<tr>
<td>-</td>
</tr>
<tr>
<td>-</td>
</tr>
<tr>
<td>Undre et al. (2006)</td>
</tr>
<tr>
<td>-</td>
</tr>
<tr>
<td>+</td>
</tr>
<tr>
<td>-</td>
</tr>
<tr>
<td>-</td>
</tr>
<tr>
<td>-</td>
</tr>
<tr>
<td>Undre et al.</td>
</tr>
<tr>
<td>-</td>
</tr>
<tr>
<td>+</td>
</tr>
<tr>
<td>-</td>
</tr>
</tbody>
</table>

Validity evidence following the unitary framework of construct validity

<table>
<thead>
<tr>
<th></th>
<th>Content</th>
<th>Response process</th>
<th>Internal structure</th>
<th>Relationship to other variables</th>
<th>Consequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healy et al. (2004)</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Undre et al. (2006)</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Undre et al.</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
1.3 Summary and Statement of the Problem

Nontechnical skills were first described in aviation as a set of attitudes and behaviours of flight crews that improved performance and safety or, if lacking, were associated with poor performance, incidents, and adverse events. Research in aviation demonstrated the trainability of these skills, leading to regulations mandating the training and assessment of CRM- or nontechnical skills of flight crews. Similarly, empirical evidence in healthcare led to the identification of OR teams’ and surgeons’ skills and behaviours that were critical for intraoperative patient safety. Reports of unacceptably high rates of preventable adverse events in first-world health care systems directed the attention to the role of human factors that were identified as root-causes in over two thirds of these events. Investigations into the role of human factors, or nontechnical skills, in patient safety culminated in the 1999 landmark report “To err is human” that underscored the role of human error in the causation of adverse events and emphasized the importance of nontechnical skills in preventing, trapping and mitigating error and adverse events. The report sparked a multitude of projects and initiatives aimed at reducing error and improving quality and safety in the OR by using aviation-style CRM-training to improve nontechnical skills.

<table>
<thead>
<tr>
<th></th>
<th>+</th>
<th>+</th>
<th>-</th>
<th>+</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sevdalis et al. (2009)</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hull et al. (2011)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Russ et al. (2012)</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 7: Validity evidence supporting scores generated with the OTAS rating system.
On a team level, large-scale initiatives reported improvements in observed nontechnical skills as well as positive effects on measures of efficiency and patient safety, including a significant reduction in perioperative morbidity and mortality. Tools and techniques to enhance OR teams’ nontechnical skills, such as preoperative team briefings, have since become routine practice in the majority of North American hospitals. Consequently, regulatory bodies for accreditation and certification across North America, the U.K., and parts of continental Europe and Australasia have mandated the integration of nontechnical skills into postgraduate curricula in the surgical specialties.

Although evidence from research in aviation and studies involving OR teams indicates that nontechnical skills are highly trainable, and several methods have been shown to be effective in teaching these skills, the influence of targeted training on the development of nontechnical skills during residency is still poorly understood. Interventions aimed at nontechnical skills of surgical residents have reported some positive short- and medium term results, but the evidence is relatively weak due to limitations in the design of many of these studies. Questions regarding when, and how best to train these skills in residency remain unanswered.

As much of learning in residency training can be explained by experiential learning theory, unguided skill development may lead to inconsistent results, since in a workplace-based environment such as the OR multiple factors may influence effective learning of nontechnical skills, both positively and negatively. Extrinsic factors include the formal and informal curriculum, as well as hidden and null curricula, while intrinsic factors comprise attitudes and personality traits. These factors will need to be considered when devising training strategies in competency-orientated training programs.

Finally, the assessment of residents’ nontechnical performance is an essential component of nontechnical skills education, allowing both the documentation of adequate skill development and the identification of deficiencies in residents’ performance. With the current move of North American residency programs toward competency-based education, summative assessments of nontechnical skills may gain importance to ensure the attainment of required competencies upon graduation. To date, however, a widely accepted approach to
the training and in-training assessment of residents’ nontechnical performance in the OR is lacking.
2

AIMS, OBJECTIVES, AND HYPOTHESES

2.1 Aims and Objectives

The aim of the present work was to develop a structured approach to integrate nontechnical skills in postgraduate surgical education, and to examine individual factors that determine innate nontechnical skills and may influence skill acquisition during surgical residency training.

The first step towards understanding the development of nontechnical skills and structuring nontechnical education in surgical residency training is to explore and characterize the current state of surgical postgraduate education with regards to these skills. Before new curricula can be developed and implemented, factors influencing the development of nontechnical skills during residency should be considered. Potential factors influencing the acquisition and development of nontechnical competencies during surgical residency training are residents’ innate nontechnical skills when entering postgraduate training; attitudes towards nontechnical skills as a measure of residents’ willingness to learn these concepts and change behaviours accordingly; and residents’ personality.

The strategy for postgraduate nontechnical skills education proposed in the present work is based on regulations and well-tried training programs in aviation, as well as theories of learning and adult education, and comprises two central elements: a mandatory basic curriculum to teach essential nontechnical knowledge, skills, attitudes and behaviours to surgical residents when they enter the residency training program; and the recurrent formative assessment and teaching of nontechnical skills during residency through feedback and debriefing during routine practice in the OR. The integration of nontechnical skills assessment in the routine workflow in the OR requires a suitable assessment tool that is aimed at resident performance in the OR and includes all relevant nontechnical skills and
competencies that are required by regulatory bodies.

Consequently, the objectives of the present work were:

**Objective 1**: To establish the current situation of nontechnical skills education among surgical residency programs in North America, including prevalence of nontechnical curricula, methods used for training and assessment, and program directors’ opinions on the design of targeted curricula.

**Objective 2**: To develop an evidence-based and reliable tool for the feasible in-training assessment of residents’ nontechnical performance in the OR.

**Objective 3**: To determine “innate” nontechnical skills of incoming surgical residents.

**Objective 4**: To evaluate the relationship between innate nontechnical skills of incoming surgical residents and personality traits on the five-factor model of human personality.

**Objective 5**: To evaluate the effectiveness of structured curricular training covering the core aspects of nontechnical skills in the first year of surgical residency training.

**Objective 6**: To evaluate the effectiveness of feedback and debriefing as instructional methods to improve nontechnical performance of surgical residents in the OR.
2.2 Hypotheses

2.2.1 Null Hypotheses

I. There is no linear relationship between personality traits on the NEO-FFI personality inventory and nontechnical skills of junior surgical trainees as assessed in standardized OR crisis simulations.

II. There is no difference in nontechnical skill development during the first 6 months of residency between residents undergoing conventional residency training and residents receiving structured training about basic nontechnical skills in a two-month curriculum concurrent to conventional residency training.

III. One session of individual structured feedback and debriefing on nontechnical skills will have no measurable effect on nontechnical performance of senior surgical trainees observed by blinded observers in the real OR.

2.2.2 Alternative Hypotheses

I. a) There is a linear, positive relationship between one or more traits on the NEO-FFI personality inventory and nontechnical skills of junior surgical trainees assessed in standardized OR crisis simulations.

I. b) There is a linear, negative relationship between one or more traits on the NEO-FFI personality inventory and nontechnical skills of junior surgical trainees assessed in standardized OR crisis simulations.

II. a) Residents who receive structured training of basic nontechnical skills in a two-month curriculum will develop superior nontechnical skills within the first 6 months of residency training when compared with residents undergoing conventional residency training.

II. b) Residents who receive structured training of basic nontechnical skills in a two-month curriculum will develop inferior nontechnical skills within the first 6 months of residency training when compared with residents undergoing conventional residency training.
III. a) Nontechnical performance of senior surgical trainees observed by blinded observers in the real OR will improve after one session of individual structured feedback and debriefing on nontechnical skills.

III. b) Nontechnical performance of senior surgical trainees observed by blinded observers in the real OR will deteriorate after one session of individual structured feedback and debriefing on nontechnical skills.
This chapter describes the needs analysis underlying the present work, with the purpose of establishing the prevalence of curricula and training interventions to teach nontechnical skills in North American residency programs.

The text of chapter 3 was reprinted with permission from Elsevier Inc. from the Journal of Surgical Education, Volume 70 (5), Dedy NJ, Zevin B, Bonrath EM, Grantcharov TP. Current concepts of team training in surgical residency: a survey of North American program directors, p. 579-584.

© 2013 Association of Program Directors in Surgery. With permission from Elsevier Inc.
3.1 Abstract

**Objectives:** The purpose of the present survey was to (1) establish the prevalence of CRM- and team-training interventions among General Surgery residency programs of the U.S. and Canada; (2) to characterize current approaches to training and assessment of nontechnical skills; and (3) to inquire about Program Directors’ (PDs’) recommendations for future curricula in graduate medical education.

**Design:** An online questionnaire was developed by the authors and distributed via email to the directors of all accredited General Surgery residency programs across the U.S. and Canada. After 3 email reminders, paper versions were sent to all non-responders.

**Participants & Setting:** PDs of accredited General Surgery residency programs in the U.S. and Canada.

**Results:** One hundred and twenty (47 percent) PDs from the U.S. and 9 (53 percent) from Canada responded to the survey. Of all respondents, 32 percent \((n=40)\) indicated conducting designated team-training interventions for residents. Three main instructional strategies were identified: combined approaches using simulation and didactic methods (42 percent, \(n=16\)); predominantly simulation-based approaches (37 percent, \(n=14\)); and didactic approaches (21 percent, \(n=8\)). Correspondingly, 83 percent \((n=93)\) of respondents recommended a combination of didactic methods and opportunities for practice for future curricula. A high agreement between responding PDs was shown regarding learning objectives for a proposed team-based training curriculum \((\alpha=0.95)\).

**Conclusions:** The self-reported prevalence of designated CRM- and team-training interventions among responding surgical residency programs was low. For the design of future curricula, the vast majority of responding PDs advocated for the combination of interactive didactic methods and opportunities for practice.
3.2 Introduction

Teamwork and communication in the OR play a key role in the delivery of safe and efficient patient care (K. Catchpole, Mishra, Handa, & McCulloch, 2008; Greenberg et al., 2007; Halverson et al., 2011; Mazzocco et al., 2009; Nundy et al., 2008). As a result, multiple initiatives have targeted teamwork, communication, and other nontechnical skills of multiprofessional teams in acute-care settings of healthcare (Buljac-Samardzic et al., 2010; Gordon et al., 2012; McCulloch et al., 2011; Weaver, Lyons, et al., 2010). In this setting, one of the most commonly used instructional strategies is CRM-training, a strategy that originated from the aviation industry and aims at enhancing team coordination and performance (Helmreich et al., 1999; McCulloch et al., 2011; Salas et al., 2008). Previous interventions based on the principles of CRM-training include large-scale initiatives such as the TeamSTEPPS® (Alonso et al., 2006) program of the Department of Defense and the Agency for Healthcare Research and Quality, and the Veteran Health Administration’s Medical Team Training (Dunn et al., 2007) program. Multiple CRM-based team-training interventions have reported very encouraging results, including a decline in observed communication failures in the OR (Halverson et al., 2011), improvements in observed teamwork (McCulloch et al., 2009; Weaver, Rosen, et al., 2010), and positive medium-term organizational outcomes such as decreases in surgical morbidity (Young-Xu et al., 2011) and mortality (Neily et al., 2010).

In addition to initiatives that were predominantly targeted at multiprofessional teams, a growing number of studies have recently reported on interventions specifically designed to teach teamwork and communication skills to surgical residents (Koutantji et al., 2008; Larkin et al., 2010; Moulton et al., 2009). At the same time, regulatory bodies for accreditation and certification have emphasized the need to teach communication and interpersonal skills in graduate medical education (ACGME, 2008). However, while multiple evidence-based methods and tools for the training of team-based skills are available to surgical educators (Agency for Healthcare Research and Quality, n.d.; ACS/APDS, n.d.), it is unclear to what extent team-based training has been implemented in surgical residency programs and what methods are being used.
The objective of the present study was to establish the prevalence and characteristics of designated team-training interventions among accredited general surgery residency programs in the U.S. and Canada. In addition, Program Directors (PD) were asked for their opinion on the design of future team training and non-technical skills curricula in graduate medical education.

3.3 Methods

**Generation of Questionnaire Items:** Questionnaire items were generated using recent literature reviews (McCulloch et al., 2011; Weaver, Lyons, et al., 2010), as well as published frameworks of teamwork (Salas et al., 2005) and non-technical skills of surgeons (Yule, Flin, Paterson-Brown, Maran, et al., 2006). The survey was designed to address two main domains: the prevalence and characteristics of current CRM- and team-training interventions, and PDs’ opinions and recommendations on the design of future curricula for non-technical skills training in surgery. A multiple-choice selection format was used for questions pertaining to the timing of training interventions, as well as methods for training and evaluation. Perceived importance of training content was assessed on a 5-point Likert-type scale with 1 being the lowest, and 5 the highest rating possible. Free-text fields were provided at the end of each question to allow for comments and suggestions. An online version of the survey was created using a web-based software (www.http:// surveymonkey.com). Prior to administration, the survey was pre-tested by two faculty members, two fellows, and two surgical residents to ensure the items were clear and unambiguous.

**Administration of the Survey:** An email containing an explanation of the purpose of the survey and a link to the online platform was sent to the directors of all accredited general surgery residency programs in the U.S. (n=253) and Canada (n=17) in the first week of April 2012. The respondents had the option to reveal their identity, or reply anonymously. Three email reminders to complete the survey were sent in weekly intervals to any
participants that had not responded. Paper versions of the survey were mailed to all Program Directors not identified as respondents of the online version. Finally, a last email reminder was sent to all individuals not identified as respondents at 10 weeks after the launch of the survey. In this email, besides a final request for participation, the PDs were asked to respond via blank email if they found the topic of the survey irrelevant. The data collection was completed by the end of June 2012.

Statistics: Descriptive statistics and frequency distributions were computed for all responses pertaining to training and evaluation methods. Agreement between respondents was computed using Cronbach’s alpha. The Statistical Package for Social Sciences software (SPSS v. 13, IBM Corporation, Armonk, NY) was used for all statistical analyses.

3.4 Results

The overall response rate was 48 percent (n=129), with 47 percent (n=120) of U.S. and 53 percent (n=9) of Canadian PDs answering the survey. Of all respondents, 66 percent (n = 85) completed the online version, and 34 percent (n = 44) the paper version of the survey. No “blank email” responses were received to the question in the final email reminder whether the PD considered the survey topic irrelevant.

Current Concepts among responding Residency Programs: Information regarding current CRM- and team-training programs was provided by 125 (97 percent) respondents. Of these, 40 (32 percent) indicated the presence of designated interventions for team training, and the vast majority (n=32, 84 percent) conducted these interventions in the first postgraduate year (PGY). Less frequently, interventions were offered in PGY 2 (11 percent, n=4), PGY 3 (3 percent, n=1), and PGY 6 (3 percent, n=1). Two respondents did not indicate the timing of initial training.
Of all programs conducting team training, 38 PDs provided detailed information on methods used for training and evaluation (Figure 7). With regards to instructional strategies, three main themes were identified: combined approaches comprising of didactic methods and simulation (42 percent, \( n = 16 \)); predominantly simulation-based approaches (37 percent, \( n = 14 \)); and didactic teaching approaches (21 percent, \( n = 8 \)). Formal training evaluation was conducted by 25 (66 percent) out of 38 programs, most frequently comprising end-of-course critiques (64 percent, \( n = 16 \)) and observations during simulation sessions (44 percent, \( n = 11 \)). Five respondents (23 percent) assessed trainees’ attitudes towards team-based skills, and six (27 percent) conducted knowledge tests. Four (18 percent) PDs reported observing residents in the real-life OR.

![Figure 7: Methods currently used for CRM and team training among responding programs.](image)

Multiple selections were possible. Total number of respondents = 38.
**PD recommendations for future team-training curricula for residents:** In total, 112 Program Directors completed this portion of the survey and provided their opinion on the “ideal” design of a team-training curriculum for surgical residents (Figure 8). The vast majority of PDs advocated for combined approaches comprising of didactic methods and opportunities for practice in simulation (83 percent, \(n=93\)). Less frequently, respondents recommended training approaches that were exclusively based on either simulation (9 percent, \(n=10\)), or didactic methods (8 percent, \(n=9\)). With regards to the time of initial exposure to team-training interventions, 45 percent (\(n=51\)) of all respondents advocated for the first year of residency, whereas 33 percent (\(n=37\)) recommended initial training prior to the beginning of residency. Fewer respondents suggested PGY2 (15 percent; \(n=17\)), PGY3 (5 percent; \(n=6\)), and PGY4 (2 percent; \(n=8\)).

![Image](image-url)

**Figure 8: Training methods recommended by responding PD for resident CRM curriculum.**

Multiple selections were possible. Total number of respondents = 112.
**Perceived Importance of Learning Content:** Learning objectives surveyed were: knowledge of human error; knowledge of patient safety; stress management; situation awareness; decision making; communication; leadership; and teamwork. Each of these objectives was rated as “important” or “very important” by more than 80 percent of respondents, with an overall high agreement on the topic among respondents (Cronbach’s α = 0.95).

### 3.5 Discussion

This study conducted a survey among PDs of accredited General Surgery residency program in the U.S. and Canada to establish the prevalence and characteristics of designated CRM- and team training interventions, and to collect opinions on the design of future team training and non-technical skills curricula in graduate medical education. The prevalence of designated CRM- and team-training initiatives among the responding residency programs was low, which is in contrast to recommendations by regulatory bodies. In the 2012 version of the Program Director Guide to the Common Program Requirements, the ACGME recommended specific training of communication and interpersonal skills in structured curricula, while expressly discouraging “on the job” learning of these skill sets (ACGME, 2011). Similarly, the Royal College of Physicians and Surgeons of Canada recently underscored that upon completion of the residency, surgeons are expected to be competent in the roles of communicator and collaborator (RCPSC, 2010b). A possible explanation for the low prevalence of specific team-training modules among surgical curricula could lie in the perceived need for such training among surgical educators. A recent survey of surgical subspecialty fellowship programs revealed a discrepancy in the perceived effectiveness of training between PDs and trainees (Francesca Monn et al., 2013). In that survey, PDs rated training effectiveness regarding communication skills significantly higher than trainees, whereas, trainees rated their perceived achievements in team building skills significantly lower than PDs (Francesca Monn et al., 2013). Similarly, multiple previous studies have shown a tendency of surgeons to overestimate the quality of teamwork and communication in the OR compared with other OR personnel (Flin et al., 2006; Makary et al., 2006;
Wauben et al., 2011). As a result of this inaccuracy in self-assessment, some surgeons may feel that team-training interventions are not required in residency training. Another explanation for the low rate of self-reported team training may be the lack of evidence for long-term effects of designated training on the development of team-based skills (McCulloch et al., 2011; Weaver, Lyons, et al., 2010), with resulting skepticism regarding the effectiveness of such training in surgical curricula.

**Current concepts and implications for future curricula:** Instructional strategies and methods for team training reported by respondents of the present study, as well as their opinions on the ideal design of a training curriculum reflect current concepts in the literature. Mixed approaches combining different didactic methods with practice have been used in over 80 percent of interventions identified in a recent systematic review of team training in healthcare (Weaver, Lyons, et al., 2010). Equally, the ACGME recently recommended combining didactic methods such as video review and small-group discussions with role-play exercises to teach interpersonal and communication skills in residency (ACGME, 2011). Evidence in the training and adult learning literature supports these recommendations. The use of video clips of positive and negative behaviour has been shown to be more effective in teaching behaviours than lectures or seminars alone (Baldwin, 1992). Using an interactive approach, trainees can be encouraged to analyze and rate behaviours observed in video clips (Flin, Yule, et al., 2007; Koutantji et al., 2008).

A further key element of training is practice. High fidelity simulation is an example of a well-established practice method that allows trainees to apply team-based skills and manage critical situations in a safe environment that closely resembles their daily workplace (Gettman et al., 2009; J. T. Paige, Kozmenko, Yang, Gururaja, et al., 2009; Powers et al., 2008; Undre, Koutantji, et al., 2007). The Division of Education of the ACS, together with the APDS, recently provided a comprehensive guide to the use of high fidelity simulation in the training of team-based skills, including detailed descriptions of the scenarios, setup, equipment needs, and learning objectives (ACS/APDS, n.d.).
In contrast, some controversy remains with regards to the sequence of instructional methods. While a sequence of information, demonstration, practice, and feedback is commonly recommended in the training literature (Salas et al., 2008), Zendejas and colleagues (2010) have recently reported that trainees who received a lecture after high fidelity crisis simulations scored significantly higher in a knowledge test than those who received the lecture in the beginning of the intervention. Moreover, recent evidence suggests that high fidelity simulation in crisis scenarios followed by structured debriefing sessions is effective as a standalone approach without the need for additional didactic teaching in lectures or seminars (Gettman et al., 2009; Hamilton et al., 2012; Knudson et al., 2008; J. T. Paige, Kozmenko, Yang, Gururaja, et al., 2009).

**Learning Objectives and Assessment of Proficiency:** Multiple valid frameworks of team-based and nontechnical skills exist to help educators define learning objectives (Salas et al., 2005; Yule, Flin, Paterson-Brown, Maran, et al., 2006). The ACS/APDS surgical skills curriculum guide contains a comprehensive evidence-based compilation of essential teamwork skills for residents (ACS/APDS, n.d.). However, there appears to be a knowledge gap as to how these skills should be assessed in residency programs, which is reflected by the heterogeneity of evaluation methods used to assess team-based and nontechnical skills (Gordon et al., 2012; McCulloch et al., 2011; Weaver, Lyons, et al., 2010). With the recent move toward competency-based education in surgery, there is a need for ongoing standardized assessment of interpersonal and communication skills, as well as a definition of proficiency standards (Nasca, Philibert, Brigham, & Flynn, 2012).

**Timing of Initial Training:** The question of when to first expose prospective surgeons to team training is another potential area of controversy. Regulatory bodies do not provide clear directives in this regard (ACGME, 2008); however, PDs largely agreed that first exposure to team training should occur in the first year of residency or even in medical school. Although the authors are not aware of any evidence in the literature, there are multiple theoretical considerations that support training in the first year of residency.
Residents, in contrast to medical students, are required to interact with patients, relatives, and other health care staff on a daily basis, and thus have multiple opportunities to apply newly acquired communication and interpersonal skills in the real world. Similar to the concept of the “pre-trained novice” (Van Sickle, Ritter, & Smith, 2006) in technical skills training, team training may provide junior trainees with a basic set of portable skills that can be expanded and built upon throughout the residency. Moreover, conducting team training in the beginning of the residency may raise residents’ awareness of the importance of nontechnical and team-based skills before they have been influenced by negative behaviours and attitudes of peers and role models on the job.

**Limitations**: The response rate of 48 percent could have predisposed the results of this study to a nonresponse bias that may have influenced the reliability of the estimates of the survey (Fowler, 2009), albeit recent surveys involving PDs of surgical subspecialties had similar response rates (Francesca Monn et al., 2013; Ganju et al., 2012; Karam, Pedowitz, Natividad, Murray, & Marsh, 2013). Non-response due to a lack of interest in the survey topic for instance may have introduced a selection bias resulting in an overestimation of the prevalence of respective training (Groves, Presser, & Dipko, 2004). In an attempt to further investigate this bias in the present survey, non-responders were asked in the final email reminder to indicate whether they thought the survey topic was irrelevant, but this email remained unanswered in all cases. Despite these limitations, the findings of the present survey indicate a discrepancy between program requirements regarding communication and interpersonal skills as defined by regulatory bodies, and current curricula in General Surgery residency programs.

### 3.6 Conclusion

The low prevalence of designated interventions for CRM- and team-training among general surgery residency programs suggests a lack of clarity regarding the necessity of such training, and controversy regarding its effectiveness. Prospective randomized controlled
trials are required to demonstrate the effects of team training on the development and maintenance of team-based and non-technical skills throughout surgical residency.
This chapter describes the development of an evidence-based rating tool for the in-training assessment of residents’ nontechnical performance in the OR.

The text of chapter 4 was reprinted with permission from Elsevier from *Surgery*, Volume 157 (6), Dedy NJ, Szasz P, Louridas M, Bonrath EM, Husslein H, Grantcharov TP. Objective Structured Assessment of Nontechnical Skills – Reliability of a Global Rating Scale for the In-training Assessment in the Operating Room, p. 1002-1013.

© 2015 Mosby Inc., with permission from Elsevier.
4.1 Abstract

**Background:** Nontechnical skills are critical for patient safety in the OR. As a result, regulatory bodies for accreditation and certification have mandated the integration of these competencies into postgraduate education. A generally accepted approach to the in-training assessment of nontechnical skills, however, is lacking. The goal of the present study was to develop an evidence-based and reliable tool for the in-training assessment of residents’ nontechnical performance in the OR.

**Methods:** The Objective Structured Assessment of Nontechnical Skills (OSANTS) tool was designed as a five-point global rating scale with descriptive anchors for each item, based on existing evidence-based frameworks of nontechnical skills, as well as resident training requirements. The tool was piloted on scripted videos and refined in an iterative process. The final version was used to rate residents’ performance in recorded OR crisis simulations, and during live observations in the OR.

**Results:** A total of 37 simulations and 10 live procedures were rated. Inter-rater agreement was good for total mean scores, both in simulation and in the real OR, with Intra-class Correlation Coefficients > 0.90 in all settings for average and single measures. Internal consistency of the scale was high (Cronbach’s Alpha=0.80).

**Conclusions:** The OSANTS global rating scale was developed as an evidence-based tool for the in-training assessment of residents’ nontechnical performance in the OR. Unique descriptive anchors allow for a criterion-referenced assessment of performance. Good reliability was demonstrated in different settings, supporting applications in research and education.

4.2 Introduction

Nontechnical skills such as teamwork, communication, and leadership have been widely recognized as essential competencies of surgeons, and have been associated with patient
safety in the OR (Greenberg et al., 2007; Kohn LT, 2000; Lingard et al., 2004; Mazzocco et al., 2009). As a result, regulatory bodies for accreditation and certification have mandated the inclusion of these skills into surgical postgraduate education (RCPSC, 2005; ACGME, 2013). In addition, the current move towards competency-based training and assessment in surgical postgraduate education within North America has revealed the need for valid, reliable and feasible rating instruments to assess all aspects of a trainee’s performance (Sanfey, Williams, & Dunnington, 2013).

To date, routine in-training assessment of nontechnical performance has not been implemented on a large scale within surgical postgraduate education (Dedy, Zevin, Bonrath, & Grantcharov, 2013), which may in part be owing to the lack of suitable rating tools. Most existing tools to rate nontechnical skills in the OR focus on team performance, and are, therefore, not targeted at the assessment of individuals (Hull et al., 2011; Mishra et al., 2009; Robertson et al., 2014; Sevdalis et al., 2008). The only existing framework specifically developed for the assessment of individual surgeons is the NOTSS system (Yule, Flin, Maran, Rowley, et al., 2008). However, NOTSS was not specifically designed to assess resident performance. In addition, nontechnical skills rating has been shown to require significant training in order to achieve acceptable reliability (Hull et al., 2013; Yule et al., 2009), which is associated with a substantial expenditure of time and cost, making a large-scale in-training assessment problematic or even unfeasible (Phitayakorn et al., 2014).

For a rating tool for in-training assessment of residents to achieve wide adoption it should be reliable after a limited amount of rater training and produce valid test scores (Norcini et al., 2011). In addition, it should be applicable to a range of educational settings including simulated environments and the real-life OR. Assessment using the tool in the OR should be embedded in the routine workflow and provide relevant information that can be used for formative feedback to promote learning and improvement of target skills that are lacking (Norcini et al., 2011). The objective of the present study was to develop an evidence-based and reliable tool for the feasible in-training assessment of residents’ nontechnical performance in the OR.

4.3 Methods

The present study was conducted in five steps: tool development; pilot testing and tool
refinement; formal rater training and calibration; validation in a simulated environment; validation in the OR. To evaluate scale reliability and generalizability, a data set from a subsequent application of the tool in an observational study was analyzed. Approval by the institutional ethics review board was obtained prior to the beginning of the study and subjects were included after they provided informed consent.

**Tool development:** The selection and definition of nontechnical skill items to be included in the new rating tool was based on existing evidence-based rating systems and frameworks of nontechnical skills in the OR (Henrickson Parker et al., 2013; Hull et al., 2011; Lingard, Regehr, Espin, & Whyte, 2006; Mishra et al., 2009; Robertson et al., 2014; Sevdalis et al., 2008; Yule, Flin, Maran, Rowley, et al., 2008). In addition, the Web sites of the ACGME (2013), the RCPSC (2014), the ISCP (2013c), and the RACS (2014) were reviewed to obtain an overview of training requirements for surgical trainees across all specialties. Particular emphasis was placed on skill items that were relevant to surgical education at the in-training and certification time-points and that were represented by behaviours that potentially were observable in residents.

**Pilot testing and tool refinement:** A preliminary version of the tool was pilot tested by two researchers who were senior surgical trainees involved in educational research. For this purpose, the tool was presented to the researchers, who subsequently reviewed videos of scripted scenarios in a simulated OR applying the new tool. The scripted simulations depicted a wide range of observable behaviours covering all skill items of the tool. In an iterative process, definitions and descriptive anchors for each skill item were assessed for clarity of wording, and refined to resolve ambiguities and to achieve an optimal characterization of the underlying nontechnical skill.

**Formal rater training and calibration:** Formal rater training included a two-hour session introducing or refreshing the concepts of nontechnical skills, discussing items of the tool,
and reviewing examples of observable skills pertaining to each item. The raters then independently viewed and rated 12 videos of 10 minutes in duration each that showed nonscripted crisis scenarios of surgical residents in a simulated OR. Ratings were compared in a one-hour discussion to calibrate raters to each other and to the scale anchors. The total time expenditure for rater training and calibration was approximately 6 hours.

**Validity evidence in a simulated environment:** Reliability of the tool was first tested in a simulation setting. Two raters were trained as described previously. Rater 1 (a surgeon-educator with 12 years of clinical experience) had previous experience in nontechnical skills assessment, but was not involved in the tool refinement. Rater 2 (a senior surgical trainee) had minimal previous exposure to nontechnical skills but had been involved in the tool refinement as described above. Both raters viewed a set of video recordings showing nonscripted crisis scenarios in a simulated OR, involving General Surgery trainees. All simulations had been conducted in a designated state-of-the-art simulation facility, and had been recorded by wide-angle wall-mounted cameras with audio capture through in-ceiling microphones. The videos were viewed in a predefined set order by each rater independently in a blinded fashion (i.e., raters were blinded to each other’s ratings). Using the tool, the raters assessed nontechnical performance of the trainees in the videos. It is important to note, that the recordings used for the validation study were different from the ones used in the rater training, thus raters had not previously seen any of these recordings.

**Validity evidence in the OR:** To test the applicability and reliability of the tool in the real OR environment, raters 1 and 2 conducted live-observations of surgical trainees of different training levels (postgraduate years 2 through 5) during full-length procedures in a General Surgery OR. Residents observed in the real OR were not the same individuals as had been subjected to the simulated crises. Observations were conducted in an unobtrusive manner, with the raters quietly observing from the sidelines of the OR. Observations were focused on procedures where the resident typically performed parts of the operations under the supervision of a staff surgeon. Again, both raters were blinded to each other’s ratings.
**Relationship with other measures of nontechnical performance:** Rater 1, who had prior experience in the use of NOTSS rating system, rated all video recordings of simulations, as well as live-observations in the OR, with the NOTSS rating system in addition to the new tool.

**Internal consistency:** To further investigate scale reliability and generalizability of the tool, we included a selected data set comprising ratings of nontechnical performance obtained in an observational study involving junior residents from various surgical specialties in standardized crisis scenarios. The data set had been obtained through independent review of simulation videos by two raters, rater 1 (as described previously), and an additional rater (trained as described previously), a fellowship-trained gynecologist with no prior experience in nontechnical skills assessment, who had not been involved in tool development, and who was blinded to the purpose of the study.

**Statistical analysis:** We evaluated interobserver agreement by calculating ICCs by using a two-way mixed model for absolute agreement. ICC values are reported as average and single measures. Internal consistency of the tool was measured with Cronbach’s Alpha, using combined ratings of all simulation videos by rater 1. Item-total correlations were performed. In addition, the theoretical change in the value of Cronbach’s Alpha if one of the items was deleted was calculated for each of the scale items. Finally, a correlation statistic was performed to evaluate the relationship between scores on the new tool and NOTSS scores. All correlations were measured by Pearson’s r. Statistical analyses were conducted using SPSS® version 20 (IBM SPSS Statistics, IBM Corporation, Armonk, NY, USA).
4.4 Results

**Tool development, pilot testing and refinement**: The final version of the Objective Structured Assessment of Nontechnical Skills (OSANTS) global rating scale comprised seven skill items: *situation awareness; decision making; teamwork; communication; leading and directing; professionalism; and managing and coordinating* (Appendix 4). For each of the seven skill items a concise definition of the underlying construct was provided, as well as descriptive anchors for the lowest (“1”), middle (“3”) and highest (“5”) score on the 5-point ordinal scale. Anchors were unique to each of the seven skill items, and determined by exemplar behaviours from previously published frameworks and rating systems of nontechnical skills (Henrickson Parker et al., 2013; Hull et al., 2011; Mishra et al., 2009; Robertson et al., 2014; Sevdalis et al., 2008; Yule, Flin, Maran, Rowley, et al., 2008).

**Validity evidence in a simulated environment**: Raters 1 and 2 each viewed six simulation videos of junior General Surgery residents involving the following crisis scenarios: anaphylactic shock; tension pneumothorax; unstable bradycardia; and carbon dioxide embolism. Good inter-rater agreement was achieved for the mean total scores, using both average (ICC = 0.95) and single measures (ICC = 0.90). Similarly, five of the seven items of OSANTS achieved good agreement (ICCs, average measures = 0.79 – 1.00). One item (*professionalism*) reached moderate agreement (ICC, average measures = 0.62). For the item *communication* the ICC could not be calculated because of a lack of variance in the scores between individual observations (Table 8).

<table>
<thead>
<tr>
<th>Item</th>
<th>SA</th>
<th>DM</th>
<th>TW</th>
<th>C</th>
<th>L &amp; D</th>
<th>PRO</th>
<th>M &amp; C</th>
<th>Mean total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>P/R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2.57</td>
</tr>
<tr>
<td>P2</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4.71</td>
</tr>
<tr>
<td>P3</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4.29</td>
</tr>
<tr>
<td>P4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4.86</td>
</tr>
<tr>
<td>P5</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>4.00</td>
</tr>
<tr>
<td>P6</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>4.71</td>
</tr>
<tr>
<td>ICC</td>
<td>0.80</td>
<td>1.00</td>
<td>0.79</td>
<td>N/A</td>
<td>0.92</td>
<td>0.62</td>
<td>0.83</td>
<td>0.95</td>
</tr>
<tr>
<td>ICC (avg.)</td>
<td>0.67</td>
<td>1.00</td>
<td>0.66</td>
<td>N/A</td>
<td>0.85</td>
<td>0.44</td>
<td>0.71</td>
<td>0.90</td>
</tr>
</tbody>
</table>
Table 8: Results from raters 1 and 2 scoring surgical trainees in a simulated environment (recorded crisis simulations)

Agreement between raters was calculated by Intra-class Correlation Coefficients (average and single measures).

Avg. = average measures; C = communication, DM = decision making, ICC = Intra-class Correlation Coefficient, L&D = leading and directing, M&C = managing and coordinating, P = Participant, PRO = professionalism, R = Rater, SA = situation awareness, sgl. = single measures, TW = teamwork.

Validity evidence in the OR: Raters 1 and 2 (as described previously) conducted live-observations during ten operations performed by trainees in a General Surgery OR. Inter-rater agreement for the mean total scores was good with an ICC= 0.95 for average and 0.90 for single measures of ICC. Similarly, good agreement was achieved for five of the seven items of OSANTS with average measures ICCs between 0.75 and 0.95. Inter-rater agreement for one item (teamwork) was moderate (average measures ICC= 0.70). For one item (professionalism) the ICC again, could not be calculated because of a lack of variance in the scores, since all but one of the participants received the same score by both raters (Table 9).
Table 9: Results from both raters scoring surgical trainees during live-observations in the OR

Agreement between raters calculated by Intra-class Correlation Coefficients (average and single measures).

Avg. = average measures, C = communication, DM = decision making, ICC = Intra-class Correlation Coefficient, L&D = leading and directing, M&C = managing and coordinating, P = Procedure, PRO = professionalism, R = Rater, SA = situation awareness, sgl. = single measures, TW= teamwork.

**Relationship with other measures of nontechnical performance**

A strong positive correlation was detected between mean total NOTSS scores from rater 1 and OSANTS scores from rater 2 (who did not score NOTSS), both for simulation videos \((r = 0.97, p = 0.001, n=6)\) and live observations \((r = 0.82, p = 0.004, n=10)\).

**Internal consistency**

Additional data from ratings of 31 video recorded simulations were used to calculate inter-rater agreement. Scale reliability and item-total correlations were based on all 37 simulation
videos rated by rater 1. Agreement between raters was good for total scores of the added 31 videos with an ICC of 0.95 for average measures and 0.90 for single measures. Inter-rater agreement for the individual items was good (ICC, average measures = 0.82–0.95) for five items, and moderate for two (communication, ICC=0.71; professionalism, ICC=0.65) (Table 10).

<table>
<thead>
<tr>
<th>ICC / Item</th>
<th>SA</th>
<th>DM</th>
<th>TW</th>
<th>C</th>
<th>L &amp; D</th>
<th>PRO</th>
<th>M &amp; C</th>
<th>Total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC (average measures)</td>
<td>0.82</td>
<td>0.89</td>
<td>0.85</td>
<td>0.71</td>
<td>0.95</td>
<td>0.65</td>
<td>0.84</td>
<td>0.95</td>
</tr>
<tr>
<td>ICC (single measures)</td>
<td>0.70</td>
<td>0.80</td>
<td>0.75</td>
<td>0.55</td>
<td>0.90</td>
<td>0.48</td>
<td>0.72</td>
<td>0.90</td>
</tr>
</tbody>
</table>

**Table 10: Agreement between two raters scoring 31 surgical trainees in a simulated environment (recorded crisis simulations)**

Inter-rater agreement was calculated by Intra-class Correlation Coefficients (average and single measures).

C = communication, DM = decision making, ICC = Intra-class Correlation Coefficient, L&D = leading and directing, M&C = managing and coordinating, PRO = professionalism, SA = situation awareness, TW = teamwork.

Internal consistency of the OSANTS rating scale as calculated from all 37 videos rated by rater 1 was high (Cronbach’s Alpha = 0.80). The item-total statistic showed substantial correlations of the individual items with the total score (Pearson’s $r = 0.51 – 0.65$) for all but one item. *Professionalism* achieved only a weak correlation with the total (Pearson’s $r = 0.22$). Removal of this item would increase Cronbach’s Alpha of the total by 0.02 points, to 0.82.

4.5 Discussion

The present study describes the development of a new tool that was designed to meet the
specific needs of program directors and postgraduate surgical educators for the in-training assessment of resident nontechnical performance in the OR. Two main design characteristics set OSANTS apart from previous rating systems: First, skill items for the tool were selected specifically to be observable in surgical residents and to be relevant in an in-training assessment context. Second, the OSANTS was designed as a global rating scale with unique descriptive anchors for each individual skill item, allowing for a criterion-referenced assessment without the need to interpret behaviours, which may improve feasibility and objectivity of the assessment.

Following Messick’s unified concept of construct validity, several sources of evidence were explored to support the validity of OSANTS test scores (Ghaderi et al., 2014; Messick, 1995). The content of the OSANTS rating scale was determined on the basis of existing evidence-based frameworks of nontechnical skills, as well as resident training requirements. It can, therefore, be assumed that the tool content reflects the overall construct of nontechnical skills in the OR and is relevant in a postgraduate education context. Response process evidence was gathered through piloting and refining the tool, ensuring raters’ responses corresponded with the underlying constructs. The validity of OSANTS scores was further supported by the convergent relationship with scores on NOTSS as the current gold standard. Variability caused by rater factors within the response process was limited by the use of descriptive anchors for different levels of performance, allowing for criterion-referenced rating. Analysis of the internal structure demonstrated good internal consistency of the OSANTS scale, and more importantly, good inter-rater agreement both in a simulated OR environment and during live-observations in the OR making the tool suitable for applications within surgical education, as well as research. Furthermore, raters from different backgrounds achieved good inter-rater agreement even after limited training.

**Design of the rating scale:** The OSANTS global rating scale was designed as a 5-point ordinal scale with unique, descriptive anchors for the highest, middle, and lowest scores of each item. The 5-point scale was structured such that the anchor for the highest score (“5”) described the optimal performance, while the lowest score (“1”) represented the worst performance or the complete lack of the particular skill. The midlevel score (“3”) was
characterized by positive aspects of the performance, with only occasional deterioration of the targeted skill during the observation. The descriptive anchors were determined by observable positive and negative exemplar behaviours from evidence-based rating systems (Norcini et al., 2011; Rosen et al., 2008). The main purpose of the descriptive anchors was to provide raters with well-defined objective criteria exemplifying different levels of performance to allow for a criterion-referenced rating to potentially reduce rater error due to bias and improve reliability of the ratings (Crossley, Humphris, et al., 2002; Swing, 2002). The design feature of descriptive anchors that are unique for each individual skill item distinguishes OSANTS from previously published rating systems of surgeons’ nontechnical skills in the OR. Previous tools used ordinal scales to indicate how well a particular behaviour was exhibited (Sevdalis et al., 2008), or provided ordinal scales with generic anchors describing the effects of observed behaviours on patient safety and/or teamwork (Hull et al., 2011; Mishra et al., 2009; Robertson et al., 2014). The NOTSS system, which has previously been the only framework to rate individual surgeons’ nontechnical skills in the OR, used a categorical scale with four rating options (poor, marginal, acceptable, and good) with descriptors indicating whether the observed behaviour improved or endangered patient safety (Yule, Flin, Maran, Rowley, et al., 2008). The use of anchors that are linked to outcomes such as patient safety, however, requires raters to interpret observed behaviours, which has been shown to increase subjectivity and the risk of rater bias (Lingard et al., 2006).

Providing unique anchors for each individual item was also thought to aid raters in distinguishing between skills, and thereby increase the discriminative power of the overall assessment. Discriminative power was pointed out previously as an area of concern with global ratings, possibly due to a halo effect, such that a candidate was rated positive on all aspects of performance although only exhibiting positive behaviour in a few areas (Swing, 2002). The results of the present study are promising with regards to the discriminative power of the OSANTS scale as several participants received scores along the spectrum of the rating scale within the same observation. In summary, the design of the OSANTS as a global rating scale with descriptive anchors that are unique to each skill item sets it apart from existing rating systems of nontechnical skills. Potential advantages over existing frameworks include criterion-referenced rating without the need to interpret behaviours,
which may improve feasibility and objectivity of the assessment.

**Selection and definition of nontechnical skill items**: The selection and definition of items was guided by two main considerations: to focus specifically on behaviours and skills that can be observed in residents, and to make the rating intuitive and feasible even after limited training. To increase feasibility, each item was described with a concise definition of the underlying construct. Following recommendations on assessment, items were defined as clearly as possible, avoiding combinations of several skills in large constructs (Rosen et al., 2008; Swing, 2002). Consequently, complex constructs were condensed to the most essential observable skills and behaviours, or divided into separate items, resulting in the seven skill items that were included in the tool.

The first two items, *situation awareness* and *decision making* have been included in virtually all previous frameworks of nontechnical skills in the OR (Henrickson Parker et al., 2013; Hull et al., 2011; Mishra et al., 2009; Robertson et al., 2014; Sevdalis et al., 2008; Yule, Flin, Maran, Rowley, et al., 2008). On the basis of a widely accepted model, *situation awareness* was defined for the OSANTS tool as an active process of perceiving and gathering information from the environment, making sense of the information in the current context, and projecting the status of the system in the near future (Endsley, 1995). The relevance of situation awareness in the OR has been underscored by data from an observational study that showed a negative correlation between surgeons’ situation awareness and the rate of technical errors (Mishra et al., 2008). The definition of *decision making* was based on previous work describing surgeons’ intraoperative decision making as identifying a problem, assessing the situation, and selecting an appropriate course of action (Pauley et al., 2011). The observable behaviours of generating options and reviewing outcomes were added to the definition to facilitate the assessment and scoring of analytical decision making strategies (Mishra et al., 2009; Robertson et al., 2014; Sevdalis et al., 2008; Yule, Flin, Maran, Rowley, et al., 2008).

The third item included was *teamwork*, which has previously been among the most commonly targeted nontechnical skills in the surgical literature (Dedy, Bonrath, Zevin, & Grantcharov, 2013; McCulloch et al., 2011). Multiple behaviours, skills, and attitudes pertaining to the construct of teamwork have been described, including team leadership,
mutual performance monitoring, backup behaviour, adaptability, and team orientation (Salas et al., 2005). To facilitate the rating, the definition of teamwork for the OSANTS scale was condensed to two teamwork-related behaviours that are observable in individual surgical trainees, and have been shown to be critical for patient safety in the OR: sharing information to establish and maintain a shared understanding among team members (Mazzocco et al., 2009; Salas et al., 2005; Yule, Flin, Maran, Rowley, et al., 2008), and supporting others (Hull et al., 2011; Mishra et al., 2009; Norcini et al., 2011; Robertson et al., 2014; Sevdalis et al., 2008). Briefing and conducting the surgical time-out were included in the descriptive anchors, as they represent formal strategies to ensure a shared understanding that have been associated with patient safety and are commonly carried out by residents (Lingard et al., 2011; Mazzocco et al., 2009; Neily et al., 2010).

The fourth item, communication, has equally been identified as a critical factor in intraoperative patient safety (Kohn LT, 2000). Similarly, communication failures have been associated with perioperative complications and errors (Joint Commission, 2014; Greenberg et al., 2007). The inclusion of communication in rating systems of nontechnical skills, however, has been inconsistent (Mishra et al., 2009; Robertson et al., 2014). In one rating system communication has been combined with teamwork to form a skill category, focusing mainly on the role of communication in the exchange of information, or when establishing a shared understanding (Yule, Flin, Maran, Rowley, et al., 2008). It can be argued, however that communication is just as important for other skills such as leadership and decision making as it is for teamwork. Thus, in the OSANTS scale, communication was included as a discrete item, focusing on its basic purpose of transmitting relevant information effectively by sending clear messages, adjusting the voice volume, addressing persons directly by name, or establishing eye contact. Focusing on the technical aspects rather than the content of communication was thought to reduce rater error and increase objectivity, because it required less interpretation of observed events by raters (Lingard et al., 2006).

The final three items that were selected for the tool were leading and directing, professionalism, and managing and coordinating, which have often been subsumed under the construct of leadership in previous scales (Hull et al., 2011; Mishra et al., 2009; Sevdalis et al., 2008; Yule, Flin, Maran, Rowley, et al., 2008). Leading and directing was defined as the trainee’s willingness or ability to assume the role of the leader in the OR when operating
as primary surgeon or assisting junior trainees, and their willingness to take charge if appropriate within a situation (Henrickson Parker et al., 2013). This aspect of leadership was thought to be particularly relevant for trainees, since the presence of a staff surgeon in the OR frequently results in a reluctance of trainees to lead, even when performing an operation as the primary surgeon (Crossley et al., 2011; Yule, Flin, Maran, Youngson, et al., 2008). Knowing when and how to assume the role of the leader is an important lesson in residency training and should be actively taught and assessed, as was emphasized in a recent report identifying the transition to the role as an independent practitioner in the OR as an area of concern and controversy among residency programs (Napolitano et al., 2014). The definition of professionalism in the OSANTS was based on the ACGME (2013) common program requirements and the CanMEDS physician roles (RCPSC, 2005) comprising attitudes and behaviours such as accountability, respect towards team members and patients, maintenance of standards and adherence to best practices. This item was deemed relevant because deficiencies in professionalism have been identified as a cause of resident remediation (Yaghoubian et al., 2012). In one study, poor professionalism in residency was a predictor for future disciplinary action as physician (Papadakis, Arnold, Blank, Holmboe, & Lipner, 2008). Managing and coordinating was included as the seventh item and was defined as the surgeon’s ability to organize activities in the OR in a time efficient and effective way by delegating tasks and using all available resources to achieve goals. In summary, skill items definitions and anchors for the OSANTS tool were specifically tailored to assess resident performance, and condensed to include only the most pertinent characteristics of the targeted competencies.

**Internal structure:** Good reliability was demonstrated for total mean scores of OSANTS. Notably, high ICCs were found in all contexts evaluated, and between raters of different backgrounds, both for average and single measures. High single-measures ICCs have implications for the feasibility of resident in-training assessment as reliable assessments may be obtained by one rater, such as a staff surgeon supervising residents during routine practice in the OR.

At the level of the individual skill items, the majority achieved average measures ICCs of at least 0.7, which has been suggested as an acceptable level of agreement for the in-training
assessment and formative feedback in medical education (Downing, 2004). In the present study, good agreement between raters was demonstrated for the items *situation awareness, decision making, teamwork, leading and directing,* and *managing and coordinating.* Moderate agreement was shown for *communication* and *professionalism.* This may be attributable to a ceiling effect because the majority of participants received good and very good scores for their communication skills and for *professionalism,* resulting in a lack of variability. Moreover, the concepts of *professionalism* targeted by OSANTS are well-known elements of physician code of conduct and can, therefore, easily be portrayed in an observational setting. Routine use of the OSANTS in everyday practice by surgical faculty is more likely to result in realistic measures of a trainee’s professionalism. Notably, item-total correlation was poor for *professionalism,* which may be due to the chosen definition of *professionalism* that was based on current resident training requirements rather than frameworks of nontechnical skills. Although removal of this item from OSANTS would improve overall internal consistency, we believe that it is an important competency that needs monitoring throughout residency. On the basis of the results from the present study, we recommend reporting total or total mean scores when using OSANTS in high-stakes settings such as research studies or in-training assessments.

**Feasibility:** Feasibility must be addressed when discussing the in-training assessment of residents, because time and financial constraints, as well as workload, often limit availability of staff to evaluate residents. The design of the tool was aimed at increasing feasibility of the rating even without previous experience in nontechnical skills, which was supported by the fact that good agreement was achieved between raters from different backgrounds, including one rater with no previous exposure to the concept of nontechnical skills. Focusing on observable behaviours rather than attitudes was thought to improve feasibility of the ratings. We acknowledge the fact, that dedicating time to conduct formal observations of residents during full-length operations may be unfeasible in the majority of settings. Therefore, incorporating observations into routine practice by training surgeons to assess residents concurrent with supervision during routine cases may economize the process. In such an in-training assessment setting, supervising staff surgeons can further facilitate the assessment of a trainee’s *situation awareness* by asking questions about the patient’s history or the
blood loss during the operation, or probing the trainee’s understanding of the local anatomy as is done frequently during routine teaching in the OR. In analogy to the concept of graded responsibility, supervising surgeons can facilitate the intraoperative assessment of decision making or leading and directing by modifying the degree of guidance during operative steps performed by the trainee. The trainee will thus be afforded an opportunity to make decisions or solve problems independently, and assume the role of the leader. Thus, the OSANTS has been tailored to surgeon educators to assess trainees whilst working with them in the OR on a daily basis.

**Limitations:** This study has two main limitations. First, the observations were limited to residents from a single residency program, which may make the findings difficult to generalize. However, numerous surgical disciplines were included, supporting applicability of the tool across specialties. Moreover, the skill items and descriptors were grounded in evidence-based frameworks and rating scales of nontechnical skills that had been tested in multiple surgical specialties and training programs across the U.S. and the U.K., suggesting that the target skills can be observed across multiple residency programs regardless of geographic location. In addition, reliability of OSANTS was demonstrated in both a simulated OR environment and in the real OR, supporting its applicability across a range of educational settings. Second, the observations were performed by designated observers rather than staff surgeons supervising the trainees in the OR. Planned next steps include training faculty members from different surgical specialties to use OSANTS while supervising residents on a daily basis in the OR to determine generalizability and feasibility of the assessment in a practice-based education context.

The OSANTS global rating scale was developed as an evidence-based tool for the in-training assessment of residents’ nontechnical performance in the OR. Unique descriptive anchors allow for a criterion-referenced assessment of performance. Good reliability was demonstrated in different settings, supporting applications in research and education.
SURGEONS’ PERSONALITY CAN INFLUENCE THEIR NONTECHNICAL SKILLS IN AN INTRAOPERATIVE CRISIS

5.1 Abstract

Background: Nontechnical skills of surgeons are critical for patient safety in the OR. Little is known about surgeons’ innate nontechnical skills, and how these relate to personality. The goal of the present cross-sectional study was to evaluate the relationship between personality and nontechnical performance of surgical trainees in simulated crisis scenarios.

Method: Junior trainees from various surgical specialties at one large training program were recruited to participate in this cross-sectional study. Participating residents completed the NEO-Five Factor personality inventory (NEO-FFI), and managed a standardized crisis in a simulated OR. Heart rate during the simulation was recorded via wearable monitors. Nontechnical performance was rated by two trained raters based on video-recordings of simulations using the NOTSS and OSANTS tools. Mental strain was measured by the coefficient of heart rate variability and controlled for in partial correlations.

Results: 31 of the eligible 56 residents participated in the study. Large inter-individual differences in nontechnical performance were observed. After controlling for confounding factors in partial correlations, a robust relationship was demonstrated between trainees’ Professionalism (OSANTS) and the NEO-FFI domains Agreeableness (r=0.45, p=0.028, df=22) and Conscientiousness (r=0.45, p=0.026, df=22). Neuroticism was correlated with an element of Decision Making (NOTSS) (r=0.61, p=0.002, df=22).

Conclusions: Personality explains some variability in surgeons’ innate nontechnical skills, with possible implications for individual training needs. Resident personality traits may have implications for the development and trainability of nontechnical skills should be further investigated.
5.2 Introduction

Nontechnical skills of health care workers have been identified as critical factors to enhance patient safety in high-risk work environments such as the OR (Kohn LT, 2000). Multiple studies have highlighted the role of nontechnical skills such as communication (Lingard et al., 2004), teamwork (Mazzocco et al., 2009), and situation awareness (Mishra et al., 2008) in the delivery of safe patient care in the OR. Consequently, nontechnical skills have been made a priority in postgraduate education (RCPSC, 2005; ACGME, 2013; ISCP, 2013c; RACS, 2014). Although previous studies have investigated the effectiveness of numerous instructional approaches to improve nontechnical performance of surgical trainees (Dedy, Bonrath, et al., 2013), the literature remains inconclusive as to how residents acquire nontechnical skills, and what the optimal approach is to teach them. Moreover, intrinsic factors may determine innate nontechnical skills or influence the development of these skills. Research in aviation for instance has linked certain personality clusters of pilots to nontechnical performance in flight simulators and training outcome after targeted interventions (Chidester et al., 1991).

An established taxonomy of human personality that may lend itself to examine the relationship between surgeon personality and nontechnical skills is the five-factor model (Costa & McCrae, 1992). The five-factor model characterizes human personality on five domains: Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness (Costa & McCrae, 1992). In previous studies using the five-factor model to explore the “surgical personality” surgical trainees were found to have higher scores than the general public in Extraversion, Openness to Experience, and Conscientiousness (Horwitz et al., 2011; McGreevy & Wiebe, 2002), while scores in Neuroticism were either lower or within the population norm (McGreevy & Wiebe, 2002). Studies evaluating the relationship between personality and surgical performance, however, are largely lacking. A recent study of 83 surgical trainees found no relationship between personality traits assessed on the NEO-
Thus far, no studies have evaluated personality factors of surgeons in relation to nontechnical skills. Identifying relationships between personality and nontechnical performance may have implications for postgraduate education, as personality may influence innate nontechnical skills, as well as skill development over time and with training. Therefore, the aim of the present study was to test the hypothesis that the five-factor model of human personality can be used to predict nontechnical performance of junior surgical trainees in standardized OR crisis simulations.

5.3 Methods

**Study Design, Settings and Participants:** The present cross-sectional study was conducted during the “Surgical Prep-Camp” for incoming residents at the University of Toronto, Department of Surgery. Residents from all surgical subspecialties were eligible to participate. Exclusion criterion was ongoing treatment with medications that affected the heart rate. Institutional ethics approval had been obtained prior to study begin (reference number 30217). Informed consent was obtained from all participants.

**Outcome measures:** Primary outcome measures were participants’ scores on the NEO™-FFI-3 personality inventory (NEO-FFI), as well as scores on the NOTSS (Yule, Flin, Maran, Rowley, et al., 2008) and OSANTS (Dedy, Szasz, et al., 2015) rating scales as assessed from video recordings of standardized OR crisis simulations.

**Measure of stress during the simulations:** Since previous studies have indicated a possible negative influence of stress on nontechnical performance in OR simulations, stress was measured as a potential confounding factor and used in partial correlations (Black et al.,
Based on previous research, the coefficient of heart rate variability (C-HRV) was selected as a surrogate marker for stress (please see below) (Czyzewska, Kiczka, Czarnecki, & Pokinko, 1983; Cordula M. Wetzel et al., 2011).

**Crisis simulations:** Crisis simulations were conducted in a fully equipped simulated OR at a state-of-the-art simulation centre. Two intraoperative crisis scenarios, based on evidence in the surgical simulation literature were used in the present study: anaphylactic shock, and unstable bradycardia (Arriaga et al., 2013). Individual case stems and mock patient charts were prepared for each surgical subspecialty. Simulations were standardized, with a physician and a medical student in full OR attire playing scripted roles of the anaesthesiology resident and scrub technician. A human patient simulator (SimMan®, Laerdal Medical, Stavanger, Norway) was used as the patient, enabling the researchers to adjust vital signs (breath sounds, chest rise, palpable pulses) and monitor readings (heart rate, blood pressure, pulse oximetry and capnography) from an adjacent control room. To increase standardization and reproducibility, the simulated crisis scenarios were preprogrammed on the simulation control software (SimMan SW version 2.5.2, Laerdal Medical, Stavanger, Norway). Simulations were recorded by a wall-mounted camera and microphone allowing for high quality audio and video data for offline evaluation.

**Randomization:** Participants were advised not to disclose the content of the simulations to their peers. To further avoid any undue familiarization with the simulated scenarios, two different scenarios were used. Each participant was randomly assigned to one of the two crisis scenarios using a block randomization protocol (www.randomizer.org). Block randomization was employed to ensure equal distributions of the two scenarios among the sample of participants.

**Procedure:** Each resident completed the NEO-FFI as well as a demographics questionnaire. The participant’s heart rate was monitored throughout the simulation using a wearable
monitor (Polar Electro, Kempele, Finland). Prior to the simulation, each participant received an introduction to the simulated OR environment, and was provided with background information about the “patient” and the case they were about to begin. The heart rate monitor was started immediately before a participant entered the simulated OR. The crisis sequence was initiated when the participant started applying skin prep on the simulated patient. After completion of the simulated crisis scenario, the participant was debriefed regarding performance and crisis management by one of the researchers.

**Data analysis:** The NEO-FFI test forms were scored by a psychometrist according to the procedure described in the published manual (Costa & McCrae, 1992). T-scores were computed using normative data from an adult population sample (Costa & McCrae, 1992). By definition, T-scores have a mean value of 50 and a standard deviation of 10 (Costa & McCrae, 1992). The C-HRV, as a surrogate marker of stress, was calculated using the following formula: C-HRV = Standard Deviation of the r-r interval / mean value of r-r interval × 100 (Czyzewska et al., 1983). Two trained raters reviewed the video recordings of the crisis simulations and scored nontechnical performance of the participants using the NOTSS and OSANTS rating systems. Raters reviewed the videos independently and in random order. One rater was blinded to the purpose of the study. Raters were: a surgeon educator with 12 years of clinical experience and previous experience in the assessment of surgeons’ nontechnical performance using NOTSS and OSANTS in more than 70 simulations; and a fellowship-trained gynecologist with 7 years of clinical experience, who had received 6 hours of dedicated training in the use of the NOTSS and OSANTS rating scales, using a standardized approach (Dedy, Szasz, et al., 2015).

**Statistical evaluation:** Descriptive statistics were calculated for all outcome measures. Pearson's correlation coefficients were computed to examine the associations between trainees’ NEO-FFI personality traits, and NOTSS and OSANTS scores. Partial correlations were performed between NEO-FFI scores and NOTSS and OSANTS scores controlling for heart rate variability, age, gender, surgical specialty, and simulation scenario. The
independent-samples t-test was used between groups. NEO-FFI T-scores of the sample were compared to the standardized norm \( M=50, SD=10 \) using a one-sample t-test. The level of significance for a two-tailed test was defined as \( \alpha = 0.05 \) for all analyses. Inter-rater agreement on NOTSS and OSANTS scores was evaluated by calculating Intra-class Correlation Coefficients (ICC) using a two-way mixed model for absolute agreement. All statistical analyses were performed using IBM® SPSS® Statistics Version 22 (Armonk, New York).

5.4 Results

Demographics: Out of an eligible population of 56 junior trainees, 31 (55 percent) volunteered to participate in the study. 20 (65 percent) of the participants were males, mean age was 27.2 \( (SD=4.1) \) years. Participants were trainees in General \( (n=14) \), Orthopaedic \( (n=4) \), Plastic \( (n=3) \), Vascular \( (n=2) \), and Cardiothoracic Surgery \( (n=1) \), Neurosurgery \( (n=3) \), Urology \( (n=2) \), and Otolaryngology \( (n=2) \).

Simulation scenarios: 16 trainees were randomized to the anaphylaxis scenario, while 15 trainees completed the bradycardia scenario. No significant differences in nontechnical skills were found between the two scenarios, except for the NOTSS element Understanding Information, on which participants scored higher in the bradycardia than in the anaphylaxis scenario \( (M=6.47, SD=1.85 \text{ vs. } M=5.06, SD=1.61; t(29)=-2.260, p=0.032) \).

Personality types, nontechnical performance, and measures of stress: Mean heart rate measured during the simulated crisis scenarios was 91.4 \( (SD=11.7) \) beats per minute, the mean r-r interval was 671.9ms \( (SD=90.1) \). The mean value of C-HRV for the sample was 6.23 \( (SD=1.38) \). Mean NEO-FFI scores and T-scores are displayed in Table 11. Mean T-scores among the sample were significantly higher than the standardized norm \( (M=50, SD=10) \) for the domains Extraversion \( (M=57.06, SD=1.38; t(30)=4.04, p<0.001) \) and
Conscientiousness ($M=55.19, SD=7.01; t(30)=4.12, p<0.001$), while T-scores of Neuroticism, Openness, and Agreeableness were within the average range. Nontechnical skills in the crisis simulations as measured in scores on the NOTSS and OSANTS rating scales are displayed in Table 12. Agreement between raters was acceptable for the category and element scores of NOTSS (average measures ICC, 0.652 to 0.925) and for the items of OSANTS (average measures ICC, 0.650 to 0.948) (Dedy, Szasz, et al., 2015), therefore, aggregate scores from both raters were used in all correlations of personality and NTS. Interestingly, large inter-individual differences in nontechnical performance were noted, with aggregate total NOTSS scores ranging from 13 through 32 (range of possible scores, 8 through 32), and total OSANTS scores ranging from 29 through 70 (range of possible scores, 14 through 70).

<table>
<thead>
<tr>
<th>NEO™-FFI-3 scores</th>
<th>M</th>
<th>SD</th>
<th>Interpretation of T-scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuroticism</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw scores</td>
<td>19.81</td>
<td>5.39</td>
<td></td>
</tr>
<tr>
<td>T-scores</td>
<td>48.71</td>
<td>7.03</td>
<td>Average</td>
</tr>
<tr>
<td>Extraversion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw scores</td>
<td>32.55</td>
<td>6.09</td>
<td></td>
</tr>
<tr>
<td>T-scores</td>
<td>57.06*</td>
<td>9.73</td>
<td>High</td>
</tr>
<tr>
<td>Openness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw scores</td>
<td>29.77</td>
<td>6.16</td>
<td></td>
</tr>
<tr>
<td>T-scores</td>
<td>52.16</td>
<td>9.72</td>
<td>Average</td>
</tr>
<tr>
<td>Agreeableness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw scores</td>
<td>31.77</td>
<td>5.16</td>
<td></td>
</tr>
<tr>
<td>T-scores</td>
<td>49.58</td>
<td>8.66</td>
<td>Average</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw scores</td>
<td>35.77</td>
<td>4.49</td>
<td></td>
</tr>
</tbody>
</table>
Table 11: Trainees’ scores on the NEO-FFI.

Raw- and T-scores are shown. T-scores are calculated using the adult age normative sample data and, by definition, have a mean (M) of 50 and a standard deviation (SD) of 10. T-scores allow for the categorization of individual scores into very low (T=34 and lower), low (T=35-44), average (T=45-55), high (T=56-65), and very high scores (T=66 and higher) when compared to the general population. *= T-scores significantly higher than standardized norms (M=50, SD=10) in a one-sample t-test (α for two-sided test = 0.05).

<table>
<thead>
<tr>
<th>NOTSS scores</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Situation Awareness</td>
<td>5.71</td>
<td>1.35</td>
</tr>
<tr>
<td>SA1 – Gathering information</td>
<td>6.23</td>
<td>1.33</td>
</tr>
<tr>
<td>SA2 – Understanding information</td>
<td>5.74</td>
<td>1.84</td>
</tr>
<tr>
<td>SA3 – Anticipating future state</td>
<td>5.77</td>
<td>1.76</td>
</tr>
<tr>
<td>Decision making</td>
<td>5.84</td>
<td>1.73</td>
</tr>
<tr>
<td>DM1 – Considering options</td>
<td>5.55</td>
<td>1.52</td>
</tr>
<tr>
<td>DM2 – Selecting/communicating options</td>
<td>5.81</td>
<td>1.64</td>
</tr>
<tr>
<td>DM3 – Implementing/reviewing decision</td>
<td>5.90</td>
<td>1.68</td>
</tr>
<tr>
<td>Communication/Teamwork</td>
<td>6.23</td>
<td>1.31</td>
</tr>
<tr>
<td>CT1 – Exchanging information</td>
<td>6.23</td>
<td>1.33</td>
</tr>
<tr>
<td>CT2 – Establishing shared understanding</td>
<td>6.58</td>
<td>1.29</td>
</tr>
<tr>
<td>CT3 – Coordinating team activities</td>
<td>5.58</td>
<td>1.75</td>
</tr>
<tr>
<td>Leadership</td>
<td>5.87</td>
<td>1.28</td>
</tr>
<tr>
<td>L1 – Setting/maintaining standards</td>
<td>6.94</td>
<td>1.26</td>
</tr>
<tr>
<td>L2 – Supporting others</td>
<td>5.35</td>
<td>1.58</td>
</tr>
<tr>
<td>L3 – Coping with pressure</td>
<td>5.58</td>
<td>1.82</td>
</tr>
<tr>
<td>NOTSS aggregate total score</td>
<td>23.65</td>
<td>4.72</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OSANTS scores</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Situation Awareness</td>
<td>6.90</td>
<td>1.92</td>
</tr>
<tr>
<td>Decision Making</td>
<td>7.10</td>
<td>2.31</td>
</tr>
<tr>
<td>Category</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>------</td>
<td>-----</td>
</tr>
<tr>
<td>Teamwork</td>
<td>7.42</td>
<td>2.00</td>
</tr>
<tr>
<td>Communication</td>
<td>8.87</td>
<td>1.45</td>
</tr>
<tr>
<td>Leading / Directing</td>
<td>7.61</td>
<td>2.68</td>
</tr>
<tr>
<td>Professionalism</td>
<td>9.29</td>
<td>1.04</td>
</tr>
<tr>
<td>Managing / Coordinating</td>
<td>6.84</td>
<td>2.28</td>
</tr>
<tr>
<td>OSANTS aggregate total score</td>
<td>54.03</td>
<td>10.11</td>
</tr>
</tbody>
</table>

Table 12: Aggregate scores from both raters’ assessment of trainees’ nontechnical performance.

NOTSS scores are presented for skill categories (depicted in italics) and elements pertaining to each category. Range of possible aggregate scores: \( \text{NOTSS}=2-8; \text{OSANTS}=2-14 \). Total scores represent the sum of NOTSS category scores and OSANTS items, respectively. CT=communication and teamwork; DM=decision making; L=leadership; \( M= \) mean; NOTSS= Non-Technical Skills for Surgeons rating system; OSANTS= Objective Structured Assessment of Nontechnical Skills rating scale; SA=situation awareness; \( SD= \) standard deviation.

**Correlations between personality and nontechnical performance**: Bivariate correlations of NEO-FFI ratings and nontechnical skills scores showed significant correlations: Neuroticism was correlated with an element of Decision Making (NOTSS, DM1; \( r=0.427, p=0.016, n=31 \)); Agreeableness with an element of Communication and Teamwork (NOTSS, CT2; \( r=0.372, p=0.039, n=31 \)), Teamwork (OSANTS; \( r=0.365, p=0.043, n=31 \)), Leadership (NOTSS, L1; \( r=0.371, p=0.04, n=31 \)), Professionalism (OSANTS; \( r=0.497, p=0.004, n=31 \)), and OSANTS total scores (\( r=0.384, p=0.033, n=31 \)); Conscientiousness with an element of Leadership (NOTSS, L1; \( r=0.467, p=0.008, n=31 \)) and Professionalism (OSANTS; \( r=0.50, p=0.004, n=31 \)). The NEO-FFI domains Extraversion and Openness were not significantly correlated with nontechnical skills.

After controlling for stress, simulation scenario, surgical specialty, gender, and age as potential confounding factors in partial correlations, three associations between personality and nontechnical performance remained significant: Neuroticism and an element of Decision Making (NOTSS, DM1; \( r=0.61, p=0.002, df=22 \)); Agreeableness and
5.5 Discussion

The present study is the first to evaluate the relationship between surgical trainees’ personality and nontechnical performance in simulated OR crisis situations. The study hypothesis was confirmed. After controlling for confounding factors, significant correlations were found for Agreeableness and Conscientiousness with Professionalism on the OSANTS scale, as well as Neuroticism and Decision Making in NOTSS. The findings have potential implications for postgraduate education, as personality may influence both natural development and trainability of nontechnical skills during surgical training.

Surgical residents’ personality: Based on the NEO-FFI T-scores, residents in the sample scored significantly higher in Extraversion and Conscientiousness than the standardized norms, while ranging average in all other traits. The findings are largely in agreement with previous investigations of surgical personality (Horwitz et al., 2011; McGreevy & Wiebe, 2002; Rosenthal et al., 2013). The literature is inconsistent with regards to Neuroticism in surgeons, with two studies reporting lower scores than the general population (McGreevy & Wiebe, 2002; Rosenthal et al., 2013), while one study found no difference to the population mean (Horwitz et al., 2011). In the present sample, Neuroticism T-scores ranged at the lower end of the average. The finding of a “surgical personality” may be due to the selection process in surgical residency programs, as well as self-selection into the specialty domain of surgery.

Relationship of personality and nontechnical performance: Personality has been defined as “the unique psychological qualities of an individual that influence a variety of
behavioural patterns (both overt and covert) across different situations and over time” ("Personality," n.d.). As a result, personality traits have been used to predict performance in various contexts (Chidester et al., 1991; Costa & McCrae, 1992). In the present study, the domains Agreeableness and Conscientiousness were found to be significant, independent predictors of residents’ professionalism during the crisis simulations. Professionalism is defined in the OSANTS as a surgeon’s commitment to patient care, their accountability, and strict adherence to standards and ethics (Dedy, Szasz, et al., 2015). Some of these attributes are consistent with traits of Conscientiousness, as conscientious individuals are thought to be hardworking, responsible, organized, punctual and reliable (Costa & McCrae, 1992). The relevance of these findings is underscored by a recent report identifying deficiencies in residents’ professionalism as a frequent cause for performance problems (Williams, Roberts, Schwind, & Dunnington, 2009).

The observed association between Professionalism and Agreeableness may be owing to the interpersonal aspects of Professionalism as defined in the OSANTS, notably respect for patients and team members (Dedy, Szasz, et al., 2015). Accordingly, Agreeableness includes traits such as altruism and cooperation (Costa & McCrae, 1992). In bivariate correlations, both Agreeableness and Conscientiousness were also associated with the NOTSS element Setting and Maintaining Standards, which describes aspects of professionalism, however this correlation was no longer significant in partial correlations. Traits of Conscientiousness and Agreeableness, as described above, reflect expectations of physicians held by the general public, as well as physician attributes defined by regulatory bodies for accreditation and certification (RCPSC, 2005; ACGME, 2013).

The lack of an association between Agreeableness and measures of teamwork was surprising, since agreeable individuals are thought to strive to support others and to cooperate (Costa & McCrae, 1992). This may be owing to the definitions of teamwork in both NOTSS and OSANTS, which emphasize the exchange of information and the establishment of a shared mental model as core teamwork behaviours (Dedy, Szasz, et al., 2015; Yule, Flin, Maran, Rowley, et al., 2008). In a meta-analysis evaluating the effects of five-factor personality traits on team performance, Peeters and colleagues (Peeters et al., 2006) previously identified Agreeableness and Conscientiousness of team members as
independent predictors of overall team performance, suggesting an important role of these traits in team functioning. It should be mentioned, however, that Peeters et al. (Peeters et al., 2006) reviewed teams from areas other than healthcare, and looked at personality in terms of team composition, while in the present study personality and performance of individual surgeons were assessed.

An unexpected yet interesting finding in the present work was a robust correlation between Neuroticism and an element of Decision Making. This finding was contrary to our expectations, as individuals high in Neuroticism are thought to be less emotionally stable and to have difficulties coping in crisis situations (Costa & McCrae, 1992). One possible explanation could be the fact that ratings of Decision Making in NOTSS are based on behaviours pertaining to an analytic and team-related approach to decision making (Yule, Flin, Maran, Rowley, et al., 2008). The element Considering Options comprises behaviours such as articulating and discussing problems, and encouraging suggestions from others (Yule, Flin, Maran, Rowley, et al., 2008). Surgeons’ intraoperative decision making strategies were recently found to be associated with individual risk tolerance: surgeons who were less willing to accept risk preferred an analytic approach, while more risk-tolerant individuals tended to make intuitive decisions (Pauley et al., 2011). Risk tolerance, particularly in threatening situations, may be influenced by personality, with surgeons higher in Neuroticism being more likely to apply risk-assessing, analytic decision making methods, thus scoring higher on the respective item of NOTSS. An alternative explanation for the positive association of Neuroticism with Decision Making could be a stress-induced increase in alertness and vigilance resulting in enhanced performance. Enhanced cognitive performance under stress can be explained by selective attention, a process where the focus of attention is narrowed to the most relevant task at hand under the influence of stress, leading to improved performance in that task (Chajut & Algom, 2003). This effect may have been more pronounced in the more neurotic and thus stress-vulnerable individuals. However, as the effect remained significant even after accounting for stress as a potential confounding factor, by controlling for C-HRV in partial correlations, this explanation is less plausible.
A further notable finding was the observed wide range of scores on nontechnical skills rating scales amongst the sample of residents, indicating large inter-individual differences in innate skills at the time of entry in the training program. Inter-individual differences in innate skills may have implications for the design of postgraduate training curricula. For instance, basic nontechnical skills training may be necessary to ensure a minimum level of nontechnical performance among incoming residents. Furthermore, trainees with deficiencies in Decision Making will likely have specific training needs that differ from those of individuals lacking Professionalism. Low Extraversion may be an indicator for individual training needs with regards to assertiveness, for instance when leading a team in the OR (Costa & McCrae, 1992). Evidence from aviation suggests that a small percentage of individuals will always reject the concepts of nontechnical skills and remain refractory to interventions (Chidester et al., 1991; Helmreich & Wilhelm, 1991). Whether personality traits can be used to predict trainability, or the optimal instructional approach to address training needs, are important questions that remain to be answered in future studies.

**Strengths and limitations:** Strengths of the study include the use of standardized simulations; inclusion of various surgical specialties improving generalizability; and reliability of nontechnical skills assessments. Three main limitations were identified: First, the recruitment of only half of the eligible population may have introduced a selection bias, with a possible overrepresentation of extraverted individuals, as introverts may have avoided participation. The recruitment process, however, was such that individuals were approached by study personnel based on availability during the surgical Prep-Camp and non-participation was largely the result of conflicting scheduling of mandatory curricular components. Second, the sample size was relatively small, which may have resulted in a lack of significant correlations between some of the variables. Although, the training program from which participants were recruited is one of the largest in North America, the number of incoming trainees available was limited. Last, performance was assessed in a simulated environment rather than the real OR. The simulated OR allows for standardized test conditions, while eliminating variability from different team members and patients.
Moreover, stressful crisis situations are seldom in the real-life OR and their occurrence is unpredictable, while the artificial introduction of a crisis would be unethical.

**Unanswered questions and future research:** Substantial inter-individual differences exist in innate nontechnical skills, with potential implications for tailoring and economizing training interventions as not all trainees may require the same amount of instruction and training. As some of these skills are associated with personality, personality testing may aid in identifying specific training needs. Whether personality can be used to select the optimal instructional approach to address these needs, or to predict the development of skills after targeted training are important questions that remain to be answered in future studies.
6

STRUCTURED TRAINING TO IMPROVE NONTECHNICAL PERFORMANCE OF JUNIOR SURGICAL RESIDENTS IN THE OPERATING ROOM: A RANDOMIZED CONTROLLED TRIAL

This chapter describes the development and evaluation of an evidence-based simulation-enhanced curriculum to provide junior surgical trainees with a basic understanding and a set of nontechnical skills relevant to the OR environment.

The text of chapter 7 was reprinted with permission from Lippincott Williams & Wilkins from Annals of Surgery, 2015 March 13, Dedy NJ, Bonrath EM, Ahmed N, Grantcharov TP. Structured training to improve nontechnical performance of junior surgical residents in the operating room: a randomized controlled trial. [Epub ahead of print].

© 2015 by Wolter Kluwer Health, Lippincott Williams & Wilkins
6.1 Abstract

Objective: The objective of the study was to evaluate the effectiveness of structured training on junior trainees’ nontechnical performance in an OR environment.

Background: Nontechnical skills have been identified as critical competencies of surgeons in the OR, and regulatory bodies have mandated their integration in postgraduate surgical curricula. Strong evidence supporting the effectiveness of curricular nontechnical skills training, however, is lacking.

Methods: Junior surgical residents were randomized to receive either conventional residency training, or additional nontechnical skills training in a two-month curriculum. Learning was assessed through a knowledge quiz and an attitudes survey. Nontechnical performance was evaluated by blinded assessment of standardized OR crisis simulations at baseline (BL) and post-training (PT) using the NOTSS and the OSANTS rating systems. Results are reported as median (interquartile ranges).

Results: Of 23 participants, 22 completed BL and PT assessment. Groups were equal at BL. At PT, curriculum-trained residents (n=11) scored higher than conventionally trained (n=11) in knowledge (12(11-13) vs. 8(6-10), p<0.001), and attitudes (4.58(4.37-4.73) vs. 4.20(4.00-4.50), p=0.008) about nontechnical skills. In a simulated OR, nontechnical performance of curriculum-trained residents improved significantly from BL to PT (NOTSS: 10(7-11) vs. 13(10-15), p=0.012; OSANTS: 23(17-28) vs. 31(25-33), p=0.012), while conventionally trained residents did not improve (NOTSS, 10(10-13) vs. 11(9-14), p=1.00; OSANTS, 26(24-32) vs. 24(23-32), p=0.713).

Conclusions: The results demonstrate the effectiveness of structured curricular training in improving nontechnical performance in the first year of surgical residency, supporting routine implementation of nontechnical components in postgraduate surgical curricula.
6.2 Introduction

Nontechnical skills have been defined as “the cognitive, social and personal resource skills that complement technical skills, and contribute to safe and efficient task performance” (p.1) (Flin et al., 2008). Nontechnical skills such as communication, teamwork, and situation awareness have been identified as critical competencies of healthcare workers to enhance patient safety (Kohn LT, 2000). Regulatory bodies for accreditation and certification in North America have recognized the need to address these competencies early in residency training and have mandated their integration in postgraduate curricula (RCPSC, 2005; ACGME, 2013).

Structured training interventions addressing nontechnical skills of surgical trainees have previously been shown to lead to significant improvements in target skills, both in the context of trauma resuscitations (Hamilton et al., 2012) and surgical ward duties (Pucher et al., 2014). The main field of activity of a surgeon, however, is the OR, a high-risk work environment where lapses and failures in nontechnical performance have been associated with errors and a potential for adverse patient outcome. (Gawande et al., 2003; Greenberg et al., 2007; Lingard et al., 2004; Mazzocco et al., 2009; Mishra et al., 2008) To date, there has been a lack of high-quality evidence supporting the effectiveness of structured curricular training to improve residents’ nontechnical performance in the OR (Dedy, Bonrath, et al., 2013). Most studies reporting on structured courses or training curricula to teach residents about nontechnical skills have reported on feasibility and positive reactions from trainees (Jones et al., 2014; Koutantji et al., 2008), or documented gains in relevant knowledge (Arora, Sevdalis, et al., 2012), while failing to demonstrate effects on nontechnical performance (Koutantji et al., 2008). On the contrary, studies evaluating the impact of training on performance focused solely on the remediation of individual deficiencies observed during simulations using feedback and debriefing (Gettman et al., 2009). The latter approach may, however, be less suitable for junior trainees with minimal prior operative experience and only a rudimentary understanding of nontechnical skills specific to the OR environment. Rather, it seems desirable to provide residents with a basic skill set and knowledge base of nontechnical competencies pertinent to the OR in the beginning of their
residency training, similar to the basic technical skills taught in simulation facilities at the start of residency (Peters et al., 2004; Sonnadara et al., 2011).

Therefore, the aim of this study was to evaluate the effectiveness of structured curricular training covering the core aspects of nontechnical skills and administered in the first year of residency training to improve junior residents’ nontechnical performance in the OR.

6.3 Methods

**Trial design:** This study was designed as a randomized controlled trial with two treatment arms: an intervention group, and a conventional training group. The allocation ratio for each treatment was 1:1. The trial was registered on www.ISRCTN.org (ISRCTN76342690).

**Study participants:** All residents commencing postgraduate training at the University of Toronto General Surgery residency program in 2012 (n=14) and 2013 (n=14) were eligible to enroll in the study. Exclusion criterion was concurrent participation or enrollment in other studies or interventions targeted at nontechnical skills. Participant enrollment throughout the study period (2012-2014) is shown in Figure 9. Approval from the Institutional Research Ethics Board was obtained before the beginning of the study (REB number 12-168), and informed consent was obtained from all participants upon enrollment.
Figure 9: CONSORT diagram illustrating participant flow throughout the study period.
Randomization: Participants were randomly assigned to the two treatment arms using a block randomization plan with a 1:1 allocation ratio generated by a web-based randomization service (www.randomizer.org). Block randomization was performed to ensure equal numbers of participants in both treatment arms.

Conventional training: Participants in the conventional training group underwent residency training at the University of Toronto General Surgery residency program comprising of regular duties in the OR, participation in the on-call schedule of the affiliated teaching hospitals, and scheduled clinical teaching sessions during weekly, protected study time. The conventional training group did not receive designated nontechnical skills training during the reported study period.

Intervention: In addition to conventional residency training as described above, the intervention group received additional structured training in a nontechnical skills curriculum, which was administered over 5 course days of 105-minute duration each, distributed over a two-month period. Course days were scheduled during residents’ protected study time. The course content was based on published training requirements, frameworks and taxonomies of nontechnical skills in the OR, as well as published interventions focusing on surgeons’ nontechnical skills (RCPSC, 2005; Dedy, Bonrath, et al., 2013; ACGME, 2013; Flin, Yule, et al., 2007; Yule, Flin, Paterson-Brown, Maran, et al., 2006). Instructional methods were based on principles of adult learning (Knowles, 1990) and evidence from the surgical education literature and comprised high-fidelity simulation, video demonstrations of positive and negative behaviour examples, interactive lectures, and group discussions (Dedy, Bonrath, et al., 2013). Each course day was structured around a predefined learning objective and began with a practice component, which was followed by facilitator-led group debriefing and a didactic component focused on the day’s learning objective. The course components were administered in an interactive format, encouraging group discussions and input from the participants. Each course day was concluded with a summary of the learning objectives, encouraging participants to apply the newly acquired
knowledge and skills to their daily duties in the OR. Simulations during the practice components were conducted in simulated environments using a human patient simulator (SimMan®, Laerdal Medical, Stavanger, Norway) as the patient and simulation technicians in the scripted roles of anaesthesiology and nursing staff. An overview of the didactic content and practical components of the five course days is provided in Table 13.

<table>
<thead>
<tr>
<th>Course day</th>
<th>Didactic content</th>
<th>Practice elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction to nontechnical skills in the OR; key competencies of teamwork (e.g., shared mental models; closed-loop communication; team leadership)</td>
<td>Unannounced simulation of patient code (cardiac arrest) involving entire group, followed by facilitator-led group debriefing</td>
</tr>
<tr>
<td>2</td>
<td>Human error in health care; systems approach to error; methods of avoiding, trapping and mitigating error; role of briefing in the OR</td>
<td>Briefing exercise in simulated OR; facilitator-led group-debriefing</td>
</tr>
<tr>
<td>3</td>
<td>Situation awareness in the OR (video examples and case vignettes from the OR and from aviation); conceptual model of situation awareness</td>
<td>OR simulation involving progressive intraoperative blood loss and gradual onset of haemodynamic instability requiring situation awareness; group debriefing</td>
</tr>
<tr>
<td>4</td>
<td>Decision making strategies; decision making in crisis situations; effects of rudeness in the OR</td>
<td>Review and group discussion of scripted video clips showing various types of behaviour in the OR</td>
</tr>
<tr>
<td>5</td>
<td>Team leadership in crisis situations; different leadership styles; course wrap-up and final discussion</td>
<td>OR simulation involving leadership and interpersonal conflict; facilitator-led group debriefing</td>
</tr>
</tbody>
</table>

Table 13: Nontechnical Skills Training Curriculum as administered to the intervention group

Duration of each course day was 1 hour and 45 minutes. OR = operating room.

Primary outcome measure: Primary outcome measure was the total score on the NOTSS rating system (Yule, Flin, Maran, Rowley, et al., 2008) as assessed by blinded analysis of
video recordings obtained in standardized crisis simulations at the baseline (BL) and post-training (PT) time points.

**Secondary outcome measures:** Secondary outcome measures were: 1) the total score on the OSANTS rating scale (Dedy, Szasz, et al., 2015), as assessed by blinded analysis of video recordings described as above; 2) perceived usefulness of the training intervention as assessed by a post-course critique administered anonymously via a web-based platform (www.surveymonkey.com); 3) mean scores on the TeamSTEPPS™ Team Attitudes Questionnaire (D. P. Baker, Amodeo, Krokos, Slonim, & Herrera, 2010) (T-TAQ) administered at BL and PT; 4) total scores on a multiple-choice quiz assessing knowledge of nontechnical skills and patient safety in the OR, administered at the PT time point. Questions in the post-course critique and the multiple-choice questions for the knowledge test were piloted to ensure clarity and exclude ambiguities before selecting questions for the PT evaluation.

**Standardized crisis simulations:** Simulations were conducted at BL and PT to assess participants’ nontechnical performance in intraoperative crisis situations. Each participant from the intervention and conventional training groups completed a crisis simulation at BL and PT. Four crisis scenarios were scripted: tension pneumothorax; anaphylactic shock; carbon dioxide embolism; and unstable bradycardia. At BL and PT, each participant was randomly assigned to one of the four crisis scenarios via computer randomization (www.randomizer.org). Randomization without replacement ensured that participants could not be assigned to the same crisis scenario at BL and PT. All simulations were conducted in a fully equipped simulated OR with a functional laparoscopy unit and a human patient simulator (SimMan®, Laerdal Medical, Stavanger, Norway). Physiological parameters of the simulated patient (palpable pulses, breath sounds, heart rate, blood pressure, pulse oximeter and capnography readings) were adjusted from the control room using the Laerdal software (SimMan SW version 2.5.2, Laerdal Medical). Crisis scenarios were preprogrammed and initiated at a predefined time point from the control room to ensure
standardization of the simulations. The roles of the anaesthesiologist, scrub technician and circulating nurse were scripted and played by simulation technicians in full OR attire. To further enhance fidelity of the simulations, a patient chart was provided and participants wore full surgical attire. Immediately before the simulation, each participant received a brief introduction to the room and the functions of the simulated patient, as well as information about the operation at hand and the “patient’s” case. The simulation began with the participant entering the room. Wall-mounted cameras and ceiling mounted microphones allowed for video and audio recording of all simulations for “off-line” evaluation. After completion of the standardized crisis simulations, participants received feedback on the technical and knowledge aspects of their performance. Participants were not given any feedback on nontechnical performance.

**Sample size:** In an *a priori* power calculation based on previous data from the authors’ research group (Zevin, 2014), the minimum sample size for a predefined power of 0.8 and $\alpha = 0.05$ for a two-sided test was calculated as 11 in each group.

**Data evaluation:** A blinded rater reviewed the video recordings of standardized crisis simulations and assessed nontechnical performance of the participants on the NOTSS and OSANTS rating scales. The rater (a surgeon educator with 12 years of clinical experience) had received dedicated training in the use of both rating tools and had previously rated nontechnical performance of >100 participants in simulated crisis scenarios as well as during live observations in the OR. The rater reviewed all videos in a random order, blinded to participant group allocation and time point (BL or PT) of the simulation. On the NOTSS scale, nontechnical performance was rated in four skill categories using a four-point scale (1=poor, 2=marginal, 3=acceptable, 4=good), resulting in possible total scores ranging from 4 to 16 (Yule, Flin, Maran, Rowley, et al., 2008). The OSANTS comprised seven items of nontechnical performance which were rated on a five-point ordinal scale with descriptive anchors for the lowest (1), middle (3) and highest (5) level of performance, resulting in possible total scores ranging from 7 to 35 (Dedy, Szasz, et al., 2015).
**Statistical analysis:** All data sets were assessed for normality of the distribution by the Shapiro-Wilk test. Parametric data were analyzed using a paired-samples t-test for within-group comparisons and an independent samples t-test for between-groups comparisons. Non-parametric statistics were employed for data that were not normally distributed, using the Wilcoxon signed-rank test for within-group comparison, and the Mann-Whitney U test for comparison between groups. The Fisher’s exact test was used for the analysis of categorical data in contingency tables. The level of significance for a two-sided test was defined as $\alpha = 0.05$. Scores are reported as medians and interquartile ranges unless stated otherwise. All statistical analyses were performed using IBM® SPSS® Statistics Version 22 (Armonk, New York).

### 6.4 Results

A total of 23 General Surgery residents in their first postgraduate year were recruited and randomized to either the intervention ($n = 11$) or the conventional training group ($n = 12$). One participant in the conventional group did not complete the post-test assessment and was excluded from the analysis. 22 participants completed the study (Figure 9). No differences in demographic parameters were found between groups at BL (Table 14). The median time interval between baseline and post-test was 35 (27 – 39) weeks in the conventional training group, and 37 (32 – 39) weeks in the intervention group ($p = 0.322$). The median number of sessions attended by the participants was 4 (4 - 5). The median time interval between the last course day of the training curriculum and the post-test assessment was 4 (2 - 9) weeks.
Table 14: Participant demographics at baseline

* = analysis by Fisher’s Exact Test; ** = Analysis by Mann-Whitney U Test. Level of significance defined as \( \alpha = 0.05 \) for a two-sided test. IQR = Interquartile Range

<table>
<thead>
<tr>
<th></th>
<th>Conventional</th>
<th>Intervention</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male : Female</td>
<td>8 : 3</td>
<td>7 : 4</td>
<td>1.000*</td>
</tr>
<tr>
<td>Age at study begin</td>
<td>25 (25-28)</td>
<td>27 (26-30)</td>
<td>.191**</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous designated team training</td>
<td>0</td>
<td>1</td>
<td>1.000*</td>
</tr>
<tr>
<td>Participation in team sports</td>
<td>5</td>
<td>7</td>
<td>.669*</td>
</tr>
</tbody>
</table>

Video evaluation of nontechnical performance: The blinded rater viewed 44 de-identified video recordings in a random order and used the NOTSS and OSANTS rating scales to score nontechnical performance of the participants. There was an equal distribution of the four different crisis scenarios, both at BL and PT, with no significant difference detected between groups. Within-group comparison showed that the intervention group improved significantly from BL to PT, with significant increases in total scores of NOTSS and OSANTS, while no significant improvement was seen in the conventional training group (Table 15). Between-group comparison at the PT time point revealed higher scores in the intervention group on both NOTSS and OSANTS although this difference did not reach statistical significance. Since the scores at BL on NOTSS and OSANTS were higher in the conventional group than the intervention group, albeit not statistically significant, change scores were calculated for both groups to account for the observed difference. Change scores were normally distributed. Parametric between-group comparison of mean change scores showed significantly higher change scores in the intervention group than the control group (Table 15), both for NOTSS \( (t(20) = 3.06, p = 0.006) \) and OSANTS \( (t(20) = 3.01, p = 0.007) \). Effect sizes, calculated using Cohen’s (1992) formula for the mean difference in change scores between groups, indicated a large effect of the intervention on both NOTSS \( (d = 1.30) \) and OSANTS \( (d = 1.29) \) scores.
Knowledge of nontechnical skills and patient safety: Participants in the intervention group scored significantly higher on the knowledge quiz administered at the PT assessment than the conventional training group (Table 15).

<table>
<thead>
<tr>
<th></th>
<th>Conventional Training</th>
<th>Curriculum Training</th>
<th>P-value (between groups)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NOTSS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BL score</td>
<td>10 (10-13)</td>
<td>10 (7-11)</td>
<td>0.224</td>
</tr>
<tr>
<td>PT score</td>
<td>11 (9-14)</td>
<td>13 (10-15)</td>
<td>0.402</td>
</tr>
<tr>
<td>p-value (within group)</td>
<td>1.000</td>
<td><strong>0.012</strong>*</td>
<td></td>
</tr>
<tr>
<td>NOTSS Δ-scores; M (SD)</td>
<td>-0.09 (2.07)</td>
<td>2.91 (2.51)</td>
<td><strong>0.006</strong>*</td>
</tr>
<tr>
<td><strong>OSANTS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BL score</td>
<td>26 (24-32)</td>
<td>23 (17-28)</td>
<td>0.120</td>
</tr>
<tr>
<td>PT score</td>
<td>24 (23-32)</td>
<td>31 (25-33)</td>
<td>0.323</td>
</tr>
<tr>
<td>p-value (within group)</td>
<td>0.713</td>
<td><strong>0.012</strong>*</td>
<td></td>
</tr>
<tr>
<td>OSANTS Δ-score; M (SD)</td>
<td>-0.55 (3.01)</td>
<td>5.55 (5.99)</td>
<td><strong>0.007</strong>*</td>
</tr>
<tr>
<td><strong>NTS Knowledge Quiz (PT)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>8 (6-10)</td>
<td>12 (11-13)</td>
<td>&lt;<strong>0.001</strong>*</td>
</tr>
<tr>
<td><strong>Teamwork Attitudes Questionnaire</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BL average score</td>
<td>4.27 (4.10 – 4.40)</td>
<td>4.37 (4.27 – 4.73)</td>
<td>0.112</td>
</tr>
<tr>
<td>PT average score</td>
<td>4.20 (4.00 – 4.50)</td>
<td>4.58 (4.37 – 4.73)</td>
<td><strong>0.008</strong>*</td>
</tr>
<tr>
<td>p-value (within group)</td>
<td>0.164</td>
<td>0.086</td>
<td></td>
</tr>
</tbody>
</table>

Table 15: Scores of conventional and curriculum-trained groups at baseline and post-training.

All scores represent medians (interquartile ranges) unless stated otherwise. Non-parametric tests were employed using the Wilcoxon sign rank test for within-group, and the Mann Whitney U test for between group comparisons. The independent-samples t-test was used for parametric between-group comparison. * Statistical significance was defined by α<0.05 for a two-sided test. BL = baseline; M = mean; NOTSS = Non-Technical Skills for Surgeons rating scale; OSANTS = Objective Structured Assessment of Nontechnical Skills; PT = post-training; SD = standard deviation.
Attitudes towards teamwork: Overall, attitudes towards teamwork in the context of patient safety were very positive amongst both groups, with no significant difference between groups at BL. At the PT assessment, however, the overall agreement with the statements about teamwork and patient safety was significantly higher in the intervention group (Table 15).

Post-course critique: Nine of the participants (82 percent) who received the curriculum training completed the anonymous post-course critique. All nine respondents strongly agreed (n = 6) or agreed (n = 3) that the course was useful for a surgical resident in the first year, and strongly agreed (n = 7) or agreed (n = 2) that a basic course in nontechnical skills should be mandatory in surgical residency training. When asked whether the course should rather be offered later in residency, one respondent agreed, while the majority disagreed (n = 6) or strongly disagreed (n = 2). Participants also agreed (n = 5) or strongly agreed (n = 4) that the knowledge and skills learned in the course can be implemented to routine practice in the operating room. Nine (100 percent) respondents stated that they had modified their behaviour in the OR after completing the course. Examples of participants’ free-text responses to how they changed their behaviour in the OR are shown in Table 16. Of note, two of the nine respondents encountered resistance from coworkers in the OR, reporting impatience or reluctance to use the surgical safety checklist for briefing (n = 2), or rejection of closed-loop communication styles by nurses (n = 1).
Participants’ free-text comments to the question “have you modified the way you behave in the OR after taking part in this course?”

<table>
<thead>
<tr>
<th></th>
<th>“More consistent and thorough safety checklists. More comfortable when confronted with challenging non-technical issues. Improved communication skills - primarily closed-loop and shared mental model.”</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>“Making a point to do all the appropriate safety checks, acknowledging the patient more, making introductions, implementing better closed-loop communication, asking for things prior to needing them.”</td>
</tr>
<tr>
<td>3</td>
<td>“It allowed for formal recognition of practices that we may recognize as poor but have not yet had the experience or training to know exactly how to avoid it or come up with an effective alternative.”</td>
</tr>
<tr>
<td>4</td>
<td>“(…) I also always “speak out loud” most steps when I am the primary operator in a case to keep other members of the team aware of the progression of the operating when appropriate. I also approach the OR as a team experience now. I used to think of it as strictly a surgeon to surgeon experience but I now recognize and have greater respect for the nursing and anesthesia component to the team. I also am able to take a pause and speak about my thoughts when I am unsure of the next step in a case and I feel this has helped me learn as staff are receptive to this instead of just stopping in a case and having the staff take over they sometimes help me work through my dilemma and let me continue onwards.”</td>
</tr>
<tr>
<td>5</td>
<td>“Taking more time to go through surgical checklist. Performing closed-loop communication.”</td>
</tr>
<tr>
<td>6</td>
<td>“More conscious of non-technical skills use from myself and colleagues.”</td>
</tr>
</tbody>
</table>

Table 16: Free-text comments of participants after receiving the non-technical skills curriculum-training

6.5 Discussion

In this study we evaluated the effectiveness of a structured training curriculum to improve junior residents’ nontechnical performance in a simulated OR environment. The intervention was received well by participants, and curriculum-trained residents clearly outperformed their conventionally trained peers in knowledge of nontechnical skills, and in their attitudes towards teamwork as a means to improve patient safety. Most importantly, though, when
comparing nontechnical performance in simulated OR crisis scenarios at baseline and post-test, the curriculum-trained residents improved significantly while conventionally trained residents’ NTS did not improve despite several months of residency training. To the knowledge of the authors, this is the first randomized controlled trial using blinded assessment in a standardized test environment to evaluate the impact of structured curricular training on junior residents’ nontechnical performance. The results support the integration of structured NTS components in postgraduate surgical curricula.

**Evaluation of training efficacy:** Evaluation of the training was based on Kirkpatrick’s (Kirkpatrick & Kirkpatrick, 2006) four-level framework and comprised the first three levels: reactions, or how participants liked the training; learning, as measured by a change in knowledge and attitudes; and transfer or behavioural change (Kirkpatrick & Kirkpatrick, 2006). Reactions to the curriculum were unanimously positive with participants finding the training useful and relevant, which was in agreement with results of previous interventions (Arora, Sevdalis, et al., 2012; Hamilton et al., 2012; Jones et al., 2014; Koutantji et al., 2008). Participants’ reactions, although a subjective measure, are thought to be critical for the success of training interventions as it is unlikely for participants to learn if the training is perceived as irrelevant or useless (Kirkpatrick & Kirkpatrick, 2006). Learning was assessed both by a change in knowledge, and a positive shift in attitudes towards teamwork. Providing trainees with relevant knowledge of NTS and how they relate to patient safety and efficiency in the OR was one of the goals of the intervention, as it is thought to be a prerequisite for the development of the desired behaviours (Kirkpatrick & Kirkpatrick, 2006). Attitudes towards relevant training objectives have been proposed as indicators of learning, as positive attitudes, or a shift towards positive attitudes, indicate an individual’s motivation to change, and to adopt new concepts and behaviours (Kraiger et al., 1993). The TeamSTEPPS® Teamwork Attitudes Questionnaire was used as it has been specifically designed to capture core concepts of teamwork relating to patient safety in the OR: team structure, leadership, situation monitoring, mutual support, and communication (D. P. Baker et al., 2010). In the present study, attitudes towards concepts of teamwork were very positive both at baseline and post-test, with the vast majority of ratings above the neutral rating,
resembling results of previous studies (Arora, Sevdalis, et al., 2012; D. P. Baker et al., 2010; Koutantji et al., 2008). Although responses in attitude surveys may to a degree be biased by social desirability, the significant difference between groups at post-test likely represented a training effect that may indicate a motivation for change among curriculum-trained residents. Transfer of training, or behaviour change as a result of the training in the present work, was measured as performance in standardized crisis simulations. Of note, the crisis scenarios used for formal assessment were different from the practical exercises used during the training, regarding both context of the scenarios and setup of the simulated environment. In addition, neither group received any feedback on nontechnical performance following the baseline simulations, to avoid “training to the test” by merely instructing participants what to do different and subsequently assessing them in a similar simulation, as practiced in previous studies (Gettman et al., 2009; Hamilton et al., 2012; Koutantji et al., 2008). Hamilton et al. (2012), for instance, used video-debriefing following simulated trauma resuscitations to teach teamwork skills to junior residents. Although the authors showed improved team functioning in trauma simulations one week following the training, the lack of a control group raised questions as to whether the improvements represented true training effects or were a result of familiarization with simulated trauma resuscitations (Hamilton et al., 2012). In the present work, in addition to using different simulation formats for training and assessment, the post-test simulations were conducted several weeks after the last course day, thereby strengthening the assumption that observed improvements in performance were a result of resident learning and transfer of training, rather than a short-lived effect of formative feedback. The majority of previous studies addressing residents’ NTS in the OR conducted the training and evaluation on the same day, which makes an intervention more feasible, but raises concerns regarding sustainable effects (Arora, Sevdalis, et al., 2012; Gettman et al., 2009; Koutantji et al., 2008). In the present study, the intervention was administered in a distributed fashion over a two-month period with post-training assessment conducted at least two weeks following the last course day. Although a narrower window for the post-training assessment would have been desirable, this was not feasible due to resident scheduling. A strength of the current study, however, is that even with delayed post-training assessment (on average 38 days after the last course session), the improvement observed in the curriculum group was sustained beyond course completion.
Instructional design: The instructional design of the training intervention combined multiple evidence-based instructional methods, including high-fidelity simulation in different settings, behaviour modeling using good and poor behaviours, interactive seminars and group discussions to accommodate differences in resident learning styles and preferences (Cook et al., 2013). The instructional design, as well as the administration of the training over several weeks allowed for distributed practice of the target skills (Cook et al., 2013). The purpose of the intervention was to teach generalizable or “portable” nontechnical skills that could be applied across a wide variety of situations, including every day routine practice in the OR, as well as stressful situations and intraoperative crises. Evidence-based instructional methods were employed to facilitate generalization of the knowledge and skills, and to enhance transfer of the training (Dedy, Bonrath, et al., 2013). For instance, trainees were shown video examples of both good and bad behaviours in the OR, and were encouraged to contribute examples from their own previous experiences in the OR (Taylor et al., 2005). At the conclusion of some course days, the course instructor defined goals for participants to apply and practice learned skills “on the job”, that is, in the real OR (Taylor et al., 2005). Simulations as practice exercises were used in conjunction with other instructional methods, as simulation-enhanced training has been shown to be superior to didactic methods alone in teaching team-based skills in a surgical context (Knudson et al., 2008). Since the trial was conducted with a pragmatic attitude course days were scheduled and offered for residents during their protected study time, but participation was voluntary. Due to conflicting resident duties, therefore, not all participants were able to attend all sessions. This is a realistic reflection of how attendance rates can be expected if a curriculum is not made mandatory by the residency program.

Learning of nontechnical skills in residency training: To date, strong evidence supporting the efficacy of structured curricular training of junior surgical residents’ nontechnical skills has been lacking (Dedy, Bonrath, et al., 2013). Efficacy of curricular components, however, needs to be demonstrated as postgraduate education is costly and curriculum time is scarce.
The present study not only demonstrated the efficacy of structured nontechnical skills training, but also demonstrated a lack of improvement in nontechnical performance during several months of conventional residency training. Existing evidence regarding the “natural” development of nontechnical skills during surgical residency training has been inconsistent. Some authors have been able to show differences in nontechnical skills between different training levels with higher scores in more advanced trainees (Black et al., 2010), or demonstrated positive correlations between nontechnical skills scores and years of postgraduate training (Crossley et al., 2011), while others found no relationship between training level and nontechnical skill (Moorthy, Munz, Forrest, et al., 2006). While previous studies compared cohorts of different training levels in cross-sectional analyses, the present study demonstrates skill development within a cohort of residents who are undergoing training in the same residency program. Providing junior residents with a basic skill set and understanding of nontechnical skills may enable them to critically reflect on own performance and modify behaviour in subsequent encounters. Moreover, creating a basic knowledge and understanding of nontechnical skills may aid junior residents in identifying and critically appraising nontechnical behaviours of peers and mentors, thus preventing them from adopting poor behaviours from role models in the “hidden curriculum” (Hafferty, 1998). In absence of other forms of structured training or dedicated feedback, nontechnical aspects of resident performance are shaped by the hidden curriculum, which leaves the development of these important competencies up to chance, or may even result in the promotion of ineffective or unwanted behaviours.

**Limitations:** The present study has four limitations: First, participating residents were enrolled in a single General Surgery residency program. To generalize the findings, dedicated training interventions need to be evaluated in other residency programs involving multiple surgical subspecialties. The concepts and skills taught in the present curriculum, however, were not specific to a General Surgical context, but would also be applicable in other operative specialties. Previous studies have similarly demonstrated that nontechnical skills can be observed across various surgical subspecialties, and current evaluation frameworks such as the NOTSS have been validated to assess nontechnical skills.
independent of the surgical specialty (Crossley et al., 2011). Second, transfer of training was assessed in a simulated environment rather than the real-life OR. Using standardized simulations with randomly assigned scripted scenarios, however, allowed for a highly standardized, blinded assessment of performance. Moreover, the simulated scenarios required participants to take over the leadership role and manage the crisis, thus potentially enabling the resident to exhibit a wide variety of nontechnical skills. In contrast, in the real OR junior residents are rarely provided with opportunities to apply nontechnical skills such as leadership or decision making, as a staff person is always present (Crossley et al., 2011). Finally, although the present study demonstrated significant within-group improvements for the curriculum-trained group, significant differences in between-group total NOTSS and OSANTS scores were not observed. Change scores were thus calculated to reflect the individual’s improvement to be used as the comparative measure. Whilst the curriculum-trained group achieved a performance improvement reflecting approximately 18 percent on the NOTSS scale (16 percent on OSANTS), the conventionally trained group showed no improvement with several individuals actually scoring lower on the post-test evaluation resulting in negative change scores. Large effect sizes for the observed differences in change scores between groups further supported the effectiveness of the training intervention.

6.6 Conclusion

In summary, the present study is the first randomized controlled trial to demonstrate the effectiveness of structured curricular training to improve junior residents’ nontechnical performance in a simulated OR environment. Random allocation to training groups and blinded assessment in scripted, standardized OR simulations underscore the strength of the results. The study results and positive responses of participants, as well as the documented lack of improvement during several months of conventional residency training, support the need for routine implementation of structured nontechnical skills training components in modern postgraduate curricula.
7
IMPLEMENTATION OF AN EFFECTIVE STRATEGY FOR TEACHING NONTECHNICAL SKILLS IN THE OPERATING ROOM: A SINGLE-BLINDED NON-RANDOMIZED TRIAL

The text of chapter 7 was reprinted with permission from Lippincott Williams & Wilkins from Annals of Surgery, Dedy NJ, Fecso AB, Szasz P, Bonrath EM, Grantcharov TP. Implementation of an effective strategy for teaching nontechnical skills in the operating room: a single-blinded non-randomized trial. [Epub ahead of print].

© 2015 by Wolter Kluwer Health, Lippincott Williams & Wilkins

7.1 Abstract

Objective: To evaluate the effectiveness of debriefing and feedback on intraoperative nontechnical performance as an instructional strategy in surgical training.

Background: Regulatory authorities for accreditation in North America have included nontechnical skills, such as communication and teamwork in the competencies to be acquired by surgical residents before graduation. Concrete recommendations regarding the training and assessment of these competencies, however, are lacking.

Methods: Non-randomized, single-blinded study using an interrupted time-series design. Eleven senior surgical residents were observed during routine cases in the OR at baseline and post-training. The NOTSS rating system was used. Observers were trained in NOTSS and blinded to the study purpose. Independent of the blinded observations, a surgeon educator conducted intraoperative observations, which served as the basis for the structured
debriefing and feedback intervention. The intervention was administered to participants after a set of (blinded) baseline observations had been completed. Primary outcome was nontechnical performance in the OR as measured by the NOTSS system. Secondary outcome was perceived utility as measured by a post-training questionnaire.

**Results:** Twelve senior surgical trainees were recruited, 11 completed the study. Average NOTSS scores improved significantly from 3.2 ($SD$ 0.37) at baseline to 3.5 ($SD$ 0.43) post-training ($t(10) = -2.55, p = 0.029$). All participants felt the intervention was useful, and the majority thought that debriefing and feedback on nontechnical skills should be integrated in surgical training.

**Conclusion:** Debriefing and feedback in the OR may represent an effective strategy to ensure development of nontechnical skills in competency-based education.

7.2 Introduction

Nontechnical skills such as situation awareness, communication, teamwork and leadership have been recognized as important competencies required of surgeons in the OR (Carthey et al., 2003; Kohn LT, 2000; Mazzocco et al., 2009; Mishra et al., 2008). Consequently, regulatory bodies across North America (RCPSC, 2014; ACGME, 2013), the United Kingdom (ISCP, 2013c), and jurisdictions within Australasia (RACS, 2014) have included nontechnical skills in resident training guidelines. Clear recommendations however, from these authorities as to how nontechnical skills should be incorporated in residency training are largely lacking. Structured curricular-based teaching of nontechnical skills to provide junior residents with a basic set of knowledge and skills in the first year of residency training was shown to be effective in a recent randomized controlled trial (Dedy, Bonrath, Ahmed, & Grantcharov, 2015). Residents trained in a simulation-enhanced curriculum showed significant improvements in their nontechnical performance in standardized OR crisis simulations, while the nontechnical skills of their peers did not improve during several months of conventional residency training (Dedy, Bonrath, et al., 2015). Simulation-enhanced curricula may thus aid the novice learner in acquiring a basic understanding of
nontechnical skills, however, throughout residency training, trainees must continue to enhance their knowledge of these skills and refine performance until they graduate as competent individuals.

Thus, competency-based residency training, as it is being implemented across North America, necessitates strategies to document adequate skill development throughout residency, and to remediate deficiencies of performance, in order to ensure skill attainment upon graduation (Holmboe, Sherbino, Long, Swing, & Frank, 2010; Iobst et al., 2010). However, recurrent simulation-enhanced training throughout residency for this purpose though may not be feasible, as curriculum time is scarce and resident work-hour restrictions require the optimal use of time for learning in the OR (Chung, 2005).

Given these shortcomings of structured training in curricula, debriefing and feedback are potential instructional strategies that could be used to help integrate nontechnical skills teaching in the routine workflow of the OR and optimize residents’ educational experience. Debriefing relates to the facilitated reflection on experiences with the purpose of identifying gaps in performance, analyzing causes for these deficiencies, and making a plan to modify behaviour accordingly in future encounters (Fanning & Gaba, 2007; Rudolph et al., 2008). Feedback comprises information about positive and negative aspects of performance that is provided to a trainee to change future performance (Ende, 1983). Both feedback and debriefing based on structured assessment of residents’ performance in the OR have been described as feasible approaches to teaching nontechnical skills in postgraduate education, although data supporting their effectiveness in this context are lacking (Spanager, Dieckmann, et al., 2015; Yule, Flin, Maran, Youngson, et al., 2008). Thus, the aim of the present study was to evaluate the effectiveness of an instructional approach using debriefing and feedback to improve resident nontechnical performance in the OR.
7.3 Methods

**Study design:** This non-randomized, single-blinded trial was designed as an interrupted time series involving the blinded assessment of a cohort of residents before and after a training intervention.

**Study participants:** Senior residents (PGY 3 through 5) and fellows in General Surgery completing a four-month rotation at a single University-affiliated tertiary centre were eligible to participate in the study. Participants were identified through departmental rotation schedules and included in the study once the first observation had been scheduled. Exclusion criteria were concurrent resident participation or enrolment in other interventional studies targeted at nontechnical or technical skills. Approval from the Institutional Research Ethics Board was obtained prior to the beginning of the study (reference numbers 29352, and 13-114), and informed consent was obtained from all participants upon enrolment.

**Sample size:** The sample size calculation was based on a previous study from the author’s group (B. Zevin, 2014). For a power of 0.8 and $\alpha = 0.05$ for a 2-sided test, the required sample size was calculated as $n = 10$.

**Outcome measures:** The primary outcome measure was nontechnical performance, as assessed using the NOTSS (Yule, Flin, Maran, Rowley, et al., 2008) rating system, during operations carried out in the OR by participating residents. The secondary outcome measure was perceived utility of the debriefing as reported by the participants on a balanced five-point Likert-type scale in a post intervention critique.

**Nontechnical skills observations:** Observations of residents’ nontechnical performance in the OR were conducted by one of four trained raters using the NOTSS rating system (Yule,
Flin, Maran, Rowley, et al., 2008). Raters consisted of two senior surgical trainees, one scrub nurse, and one research coordinator. All raters had received targeted training in the use of NOTSS, and in the assessment of nontechnical performance based on published recommendations, using scripted videos and recordings of simulated crisis scenarios, as well as supervised observations in the OR (Hull et al., 2013). Raters were assigned to trainees based on availability at the time of enrolment. Once assigned to a trainee, a rater completed all observations of that trainee during the study period to minimize rater errors due to leniency or stringency. Of note, all raters were blinded to the purpose of the study, thus unaware of the presence of a training intervention, to minimize the risk of observer bias. In order to obtain reliable assessments that were representative of an individual’s performance, multiple assessments were conducted at baseline and post-training. Based on a previously published reliability model using a D-study, a target number of four assessments before and after the intervention was selected to achieve a reliability coefficient of 0.70 or more (Crossley et al., 2011). During each observation, raters scored observed performance on the four categories of NOTSS (situation awareness; decision making; communication and teamwork; and leadership) (Yule, Flin, Maran, Rowley, et al., 2008).

**Study procedure:** Upon enrolment, participating residents were informed of the purpose of the study. Residents were then observed and rated by a designated rater during full-length operations carried out, at least in part, by the trainee. Observations were conducted throughout the duration of the resident’s affiliation to a surgical team (two months) to obtain baseline and post-training measures of nontechnical performance. The raters were not aware of an intervention and had been instructed to conduct eight observations spaced out over the resident’s surgical rotation. After the baseline observations each trainee was independently observed by a surgeon-educator (a different individual from the above mentioned raters) during at least one full-length operation. Following the observation by the surgeon-educator the intervention (please see *Intervention* below) was administered. After completion of the post-training assessments, trainees were asked to complete a post-course critique.
**Intervention:** The intervention was administered by one of two surgeon-educators who met with the resident for an individual debriefing and feedback session. Both surgeon-educators had more than 10 years of clinical practice and extensive experience in assessing and debriefing nontechnical skills. The sessions were informed by a structured observation of the trainees during at least one operation performed in part by the trainee using the NOTSS framework and the OSANTS framework (Dedy, Szasz, et al., 2015). The OSANTS scale was used in addition to NOTSS, as it is focused specifically on nontechnical behaviours that can be observed in trainees (Dedy, Szasz, et al., 2015). Leadership, for instance, is defined in OSANTS as the trainee’s ability and willingness to assume the role of the leader, and their assertiveness in doing so (Dedy, Szasz, et al., 2015). Moreover, communication is a distinct item of OSANTS and defined by technical aspects (i.e., effectiveness of communicating in terms of closing the loop, speaking loud enough etc.), whereas in NOTSS it is combined with teamwork to describe the exchange of information as a process (Dedy, Szasz, et al., 2015; Yule, Flin, Maran, Rowley, et al., 2008). OSANTS was not used by blinded raters as an outcome measure, since validity evidence was still pending at the time of data collection for the present study. Following generally accepted recommendations on debriefing and feedback, the sessions were conducted at a time and location convenient to the resident (Ende, 1983; Fanning & Gaba, 2007; Rudolph et al., 2008). The intervention was largely based on a previously published four-step model of feedback and debriefing in formative assessment comprising the identification of performance gaps, provision of feedback regarding observed gaps, investigation of potential causes, and discussion of strategies to close gaps (Rudolph et al., 2008). In our approach, residents were first informed of the individual focus of the assessment by defining target nontechnical skills (e.g., situational awareness; decision making; communication; teamwork; leadership and professionalism). Residents were then encouraged to reflect on their performance in these target skills, facilitated by the surgeon-educator through the provision of structured feedback. Feedback comprised both positive and negative examples of performance, and was based solely on behaviours observed first-hand by the surgeon-educator, and limited to those behaviours that were amenable to change. The surgeon-educator then facilitated the identification of performance gaps by the trainee, which was followed by a discussion of possible changes in behaviour to improve performance (Rudolph et al., 2008). Conversely,
positive behaviours were reinforced. The sessions were concluded by the agreement of the resident and educator on a plan to change target behaviours in order to improve performance. After the session, residents received a written summary of their performance, including target skills with definitions and examples, and the planned changes in behaviour.

**Statistics:** Data were analyzed for normality of the distribution by Shapiro-Wilk’s test. NOTSS scores were calculated for each observation as the mean of the four category scores. For comparison between observations at baseline and post-training, average values of all observations at baseline and post-training were calculated. Within-group comparisons were conducted by paired-samples t-tests. The level of significance for a two-sided test was defined as $\alpha=0.05$. Scores are reported as means and standard deviations for parametric data, and medians and interquartile ranges for non-parametric data. All statistical analyses were performed using IBM® SPSS® Statistics Version 22 (Armonk, New York).

7.4 Results

**Participants:** Of the 22 eligible participants, 17 expressed interest in the study of which 12 scheduled observations of their cases with a designated observer. For one participant, post-training observations could not be obtained due to unavailability of the rater. Thus, complete data sets on 11 participants were included in the final analysis.

**Number of observations:** Overall, 69 observations of full-length procedures were conducted in the OR, comprised of 40 baseline and 29 post-training observations. The types of procedures observed depended on the training level of the participants and their current rotation and included laparoscopic hemi-colectomies, laparoscopic Roux-en-Y gastric bypass procedures, laparoscopic cholecystectomies, breast operations, and a tracheostomy. At baseline, each participant was observed during a median number of four (range, 1 – 7) cases. After the intervention a median number of three (range 1 – 4) observations were
obtained. The individual deviations from the target number of four observations at baseline and post-training were due to conflicting schedules of the residents and raters.

**Intervention**: All participating residents received one feedback/debriefing session, with the exception of one participant who received the intervention in two parts due to time constraints, with the first session on the day of the observed operation and a second session on the following day. The median duration of the intervention was 30 minutes (range, 20 – 40 minutes). Deficiencies in nontechnical performance that were addressed by the surgeon-educator in the sessions are shown in Table 17.

<table>
<thead>
<tr>
<th>Skill category</th>
<th>Examples</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication</td>
<td>Ambiguity due to not addressing people by name; problems with audibility due to not speaking loud enough; overall lack of information exchange</td>
<td>8</td>
</tr>
<tr>
<td>Teamwork (establishing a shared understanding)</td>
<td>Deficiencies in the briefing and/or surgical pause (e.g., not involving all team members; omitting important issues; failure to update team on new information)</td>
<td>6</td>
</tr>
<tr>
<td>Decision making</td>
<td>Not involving others; not asking for input from staff when struggling with task; lack of problem definition</td>
<td>5</td>
</tr>
<tr>
<td>Situation Awareness</td>
<td>Failure to anticipate equipment needs; deficiencies in perceiving cues from environment; tunnel vision</td>
<td>4</td>
</tr>
<tr>
<td>Leadership</td>
<td>Lack of assertiveness when operating as primary surgeon (e.g., remaining passive, waiting for prompts from staff)</td>
<td>4</td>
</tr>
<tr>
<td>Teamwork (coordination)</td>
<td>Addressing circulating nurse with a request, although the nurse is currently busy with a different task or not in the room</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 17: Deficiencies in nontechnical performance observed by surgeon-educators and addressed in the debriefing and feedback discussions with residents.

Changes in observed performance: Data were normally distributed. Average mean NOTSS scores improved significantly from 3.2 (SD 0.37) at baseline to 3.5 (SD 0.43) after the intervention ($t(10) = -2.55, p = 0.029$), equating to a medium-sized effect of the intervention of $d = 0.74$ according to Cohen’s (1992) formula.

Participants’ reactions to the intervention: All participants completed the evaluation form. All respondents found the intervention useful (n=6) or very useful (n=5). Ten participants felt very strongly (n=5) or strongly (n=5) that formal feedback and debriefing should be integrated in surgical residency training, one participant was undecided. Regarding the frequency of formal feedback and debriefing on NTS one resident suggested weekly sessions, while the remaining residents felt fortnightly (n=3) or monthly (n=7) sessions would be sufficient. The majority (n=10) of residents very strongly (n=5) or strongly (n=5) agreed that informal feedback and debriefing from supervising staff surgeons on nontechnical performance in the OR should be integrated in routine training. One participant felt less strongly about this. Informal feedback was suggested as daily (n=4), weekly (n=3), fortnightly (n=1), monthly (n=2), and once per rotation (n=1) sessions. When asked about resistance and barriers to implement learning points from the sessions into daily practice, the majority (n=8) of respondents denied any resistance, while three residents felt they had encountered some (n=2) or strong (n=1) resistance. The participant who indicated strong resistance toward change specified this as a disapproval of the surgical safety checklist by the staff surgeon.
7.5 Discussion

In the present study, we demonstrated the effectiveness of combined debriefing and feedback in the OR as an approach to the teaching of nontechnical skills in surgical residency training. Strengths of the study include a standardized intervention based on published evidence and current recommendations; strictly blinded assessment by trained raters; assessment of relevant nontechnical skills in the clinical workplace using a comprehensive and well-tested tool; and the assessment across a variety of different procedures. A significant improvement of residents’ nontechnical performance was observed after a single session of individual debriefing and feedback. Moreover, the intervention was well received by trainees, who felt both formal and informal feedback and debriefing on nontechnical aspects of operative performance should be integrated in surgical residency training.

The need for new training strategies: The current move to competency-based training in North America necessitates the implementation of reliable, effective and feasible strategies to ensure the development of critical nontechnical skills, and to allow for the identification and remediation of deficiencies in performance, if present (Holmboe et al., 2010; Norcini et al., 2011). Formative assessment has been defined as assessment for learning, as opposed to the summative assessment of learning (Marriott, Purdie, Crossley, & Beard, 2011). Formative assessments should be ongoing and timely, integrated in routine workflow and provide residents with specific feedback on their performance (Norcini et al., 2011). In the U.K., competency-based formative assessment was introduced in 2007 with the ISCP (Eardley, Bussey, Woodthorpe, Munsch, & Beard, 2013). Competency-based assessments within the ISCP are based predominantly on formative assessments in the workplace, (i.e. the OR) by means of Direct Observation of Procedural Skills (DOPS) during minor procedures, and Procedure-Based Assessments (PBA) during major procedures (Eardley et al., 2013). The ISCP mandates the completion of DOPS and PBAs in regular intervals throughout residency training in order to ensure predefined competency-levels have been achieved when promoting trainees to the next stage within the training program (Marriott et al., 2011). Checklist-type forms aid educators to structure the PBAs and assess trainees’
performance on relevant competencies, which are then discussed in postoperative feedback sessions (Marriott et al., 2011). Although some aspects of nontechnical performance have been integrated into PBAs (e.g., “communicates clearly and consistently with the scrub team/ anaesthesiologist” (ISCP, 2013b), the majority of task-specific and global competencies are related to technical aspects and knowledge of the procedure (ISCP, 2013b).

**Design of the intervention:** The training intervention was based on published guidelines and recommendations on debriefing and feedback. Debriefing was administered by experienced surgeon-educators at a time and place convenient to the trainee (Fanning & Gaba, 2007); learning objectives were clarified (Rudolph et al., 2008); residents were encouraged to play an active role (Rudolph et al., 2008); and trainer and trainee agreed on concrete changes in behaviour based on identified gaps (Rudolph et al., 2008). Feedback was administered timely after the observed procedure (Parmar & Delaney, 2011); scheduling was based on trainee preference (Ende, 1983); feedback was descriptive and nonjudgmental, and based on first-hand observations (Ende, 1983); positive and negative performance was discussed, providing concrete examples from observed behaviours (Ende, 1983). Of note, participants were aware of the purpose of the study to improve nontechnical skills, which may have influenced their behaviour while being observed. Informing participants about the purpose of the study, however, was an essential component of the intervention because active participation of the trainee is crucial for the effectiveness of debriefing (Fanning & Gaba, 2007; Rudolph et al., 2008). In addition, the anticipation of timely feedback itself has been shown to lead to improved performance and as such may be utilized to enhance positive effects of interventions (Kettle & Haubl, 2010).

**Learning of nontechnical skills in the OR:** Following experiential learning theory residents learn nontechnical skills by reflecting on, and making sense of experiences made in the OR, thereby transforming the experiences to knowledge (D. A. Kolb, 1984). Teunissen et al. (2007) recently proposed a theoretical model of resident learning in the
clinical environment, which reflects the experiential learning process as conceptualized by Kolb (D. A. Kolb, 1984), while emphasizing the importance of external factors such as feedback and comments from staff surgeons on resident learning (Teunissen, Scheele, et al., 2007). Debriefing and feedback represent well-tried evidence-based instructional strategies to guide experiential learning of nontechnical aspects of performance, as has been demonstrated in simulation-based education (Cook et al., 2013; Fanning & Gaba, 2007). The importance of feedback in competency-based education was recently emphasized (Holmboe et al., 2010). A growing body of evidence indicates effectiveness of debriefing and feedback on the development of technical skills in the clinical environment (i.e., the OR) (Grantcharov et al., 2007). In addition, debriefing on observed operative performance was recently proposed as a method to teach NTS to surgical trainees (Yule, Flin, Maran, Youngson, et al., 2008). Yule et al. (2008) trained staff surgeons in the use of the NOTSS framework to assess residents’ nontechnical performance and facilitate the reflection on their performance in postoperative debriefing sessions. Although the provision of feedback to residents was perceived as useful by staff surgeons, the authors reported challenges in motivating staff surgeons to complete a predefined number of debriefings (Yule, Flin, Maran, Youngson, et al., 2008). Spanager et al. (2015) recently reported positive responses from residents in a qualitative study to evaluate the feasibility of postoperative feedback on nontechnical skills using the Danish modification of NOTSS. Neither study, however, evaluated the residents’ learning or behaviour changes as a result of the intervention (Spanager, Dieckmann, et al., 2015; Yule, Flin, Maran, Youngson, et al., 2008). Effectiveness of educational activities in postgraduate surgical education should be demonstrated prior to their implementation in curricula, as curriculum time is limited and costly. Without explicit instruction, however, behaviours and attitudes pertaining to nontechnical aspects of performance are predominantly acquired through implicit teaching in the “hidden curriculum” (Hafferty, 1998). The hidden curriculum comprises expectations, values, and norms that exist in the educational environment and influence learning, both positively and negatively, which may result in unintended learning outcomes (Portelli, 1993). The perceived resistance from supervisors against the concepts taught in the debriefings, as reported by three participants in the present study, indicates a potential negative influence of the hidden curriculum. Thus, relying solely on learning of
nontechnical skills in the hidden curriculum may leave the acquisition of these important competencies to chance. The results of the present study support the use of debriefing and feedback to enhance and guide experiential learning of nontechnical skills in surgical postgraduate training.

**Limitations:** Four limitations need to be addressed: First, the number of observations of each participant deviated from the planned four observations at baseline and post-training, with potential implications for the reliability of the scores. In a study to evaluate the psychometric properties of NOTSS in assessing surgical trainees’ nontechnical performance in the OR, Crossley et al. (2011) recently demonstrated that, when using one rater per case, a minimum of four observations were required to achieve a reliability coefficient of more than 0.7. Crossley et al. (2011), however, used different and sometimes minimally trained raters for each individual case while in the present study well-trained raters performed all observations of one participant in a blinded fashion (i.e., they were unaware of the study purpose and the presence of a training intervention) (Crossley et al., 2011). This study was a pragmatic trial, and as such likely reflects the response and participation rates of an educational activity that is implemented in an educational environment on a voluntary basis. Higher compliance with observations would potentially be achieved if teaching by debriefing and feedback were made mandatory, as piloted in the ISCP (Marriott et al., 2011).

Second, observations in this study were limited to a single specialty and site. Although previous observations using NOTSS have shown that NTS in the OR remain relatively stable across different procedures and specialties (Crossley et al., 2011), further studies involving multiple sites and surgical specialties should confirm generalizability of the results.

Third, observations were carried out by trained raters attending the cases with the sole purpose of rating resident nontechnical performance. This may be unfeasible in some programs, as a recent report has emphasized the high cost and time expenditure associated with the formal training of raters for nontechnical skills assessment in residency programs.
(Phitayakorn et al., 2014). Thus, a more feasible approach may be to train staff surgeons in the assessment of nontechnical skills, and integrate assessment and teaching of these competencies into the daily workflow of the OR. In this setting staff surgeons would be able to facilitate the assessment of skills that may be difficult to observe in trainees such as leadership or decision making, by modifying the degree of active supervision and assistance. All but one resident in the present study felt strongly that frequent informal feedback on nontechnical skills from clinical supervisors should be integrated into routine residency training, which was in agreement with responses from a recent trial of nontechnical skills feedback in a Danish residency program (Spanager, Dieckmann, et al., 2015). A lack of motivation on the part of the staff surgeons may represent a barrier to the implementation, as previously reported (Yule et al., 2008). Thus, identifying a small committed group of staff members with an interest in education and nontechnical skills seems critical for the success of this strategy, and may also help to ensure reliability of ratings and consistent quality of teaching.

Lastly, the results need to be interpreted with some caution due to the lack of a control group to determine any improvements through workplace based unguided learning over the course of the rotation. Due to the limited number of residents in General Surgery at our institution, for there to have been a control group, the data collection would have required several years to achieve the predetermined sample size, introducing additional confounding factors from changes in the learning environment. In addition, in a recent randomized trial of simulation-enhanced NTS training, we demonstrated that, without targeted training, residents’ nontechnical skills did not improve over several months of residency training. (Dedy, Bonrath, et al., 2015) The lack of a control group was partially compensated for by using the participants’ performance at baseline as control in an interrupted time-series design with blinded raters.
7.6 Conclusion

Formative assessment of residents’ nontechnical performance in the OR followed by timely debriefing and feedback may represent an effective strategy to ensure the acquisition and development of nontechnical skills in competency-based surgical training.
8

GENERAL DISCUSSION

8.1 Summary of Study Findings

The aim of the present thesis was to develop a structured approach to integrate nontechnical skills in postgraduate surgical education, to evaluate the effectiveness of the main components of the proposed approach, and to explore factors that may influence the development of these skills during residency training.

After a review of the literature on human factors research in aviation and healthcare and on nontechnical skills training in surgery to identify learning objectives and instructional strategies, we conducted a survey of residency program directors as a needs analysis to establish the current situation of nontechnical skills education among surgical residency programs in North America. The results of the needs analysis, as presented in chapter three of my thesis, revealed a discrepancy between requirements by regulatory bodies and professional associations on the one hand, and the reality of teaching practice in residency programs across North America on the other hand. Of 129 program directors (48 percent) that responded to the survey, only one third indicated the presence of targeted curricula or curricular components to address nontechnical skills in postgraduate education in their program. It is possible that the true percentage of programs with nontechnical training components is even lower, since the program director who responded to the survey potentially represented those who were interested in the topic or at least open to the concepts addressed in the survey. The written response from one program director who stated that “all of this has been taught on a daily basis for a hundred years and is part of a surgical residency program...waste of time to translate this to a curriculum...must be learned on the job every day” indicated that there may be a problem with the perceived utility of nontechnical skills education as a component of formal surgical curricula. The lack of guidance from regulatory authorities regarding the training and assessment of nontechnical skills may also have contributed to the incomplete implementation in training curricula.
In chapter four I described the development of a tool for the in-training assessment of residents’ nontechnical performance in the OR. The content of the OSANTS scale was based on empirical evidence from human factors research in aviation and healthcare, existing rating systems and taxonomies of nontechnical skills, and resident training requirements. Descriptive anchors for the lowest, middle, and highest score were defined for each of the seven items of OSANTS to allow for an objective criterion-referenced assessment on the five-point rating scale. Validity of OSANTS test scores was supported by evidence based on content, response process, internal structure, and relations to other variables, such as NOTSS scores. Good reliability was demonstrated for scores generated with the tool both in a simulated environment and in the real-life OR, making OSANTS a suitable rating instrument to be used in formative and summative assessment of nontechnical performance in competency-based education.

In chapter five I described a cross-sectional study aimed at exploring intrinsic factors that determine individual training needs and potentially influence the development of nontechnical skills during residency. In particular the study was aimed at the relationship between personality traits and innate nontechnical skills of junior surgical residents. The main findings of the study were large inter-individual differences in innate nontechnical skills of junior trainees when entering residency training; and an association of some nontechnical skills and skill elements with residents’ personality traits on the five-factor model. The Null hypothesis stated under (I), chapter 2, was thus refuted. Since a significant positive correlation was found between personality and nontechnical skills, the alternative hypothesis (Ia) was confirmed. These findings may be interpreted towards a need for more individualized training, for instance through individual debriefing and feedback, and underscore the importance of assessment to determine training needs and potential deficiencies in performance. Moreover, the results indicated that personality testing could potentially be used to identify individual training needs early in residency training and select educational content accordingly.

In chapter six and seven, I presented the two main components of a proposed structured approach to nontechnical skills education in surgical residency training: initial training in a basic nontechnical skills curriculum; and recurrent training and assessment through
individual debriefing and feedback on nontechnical performance in the OR. The basic curriculum has the goal of providing junior residents with a set of fundamental nontechnical knowledge and skills relevant for the OR environment in the first year of residency training. Formative assessment of supervised practice in the OR will subsequently help to ensure adequate development of nontechnical skills throughout residency and allow for the remediation of deficiencies if detected. The effectiveness of a structured curriculum in the first year of residency was evaluated in a randomized controlled trial, as detailed in chapter six. Residents who were trained in the two-month simulation-enhanced curriculum, administered over five sessions during protected study time, showed a significant improvement in nontechnical performance in OR crisis simulations. No improvement of nontechnical performance was demonstrated in the group of residents undergoing several months of conventional residency training without targeted nontechnical components. Moreover, curriculum-trained residents outperformed their conventionally trained peers in knowledge and attitudes about nontechnical skills and patient safety in the OR. Based on these results, the Null hypothesis stated under (II) in chapter 2 was refuted. The alternative hypothesis stated under (IIa) was confirmed. The results support the integration of structured nontechnical skills training in a formal curriculum in the beginning of surgical residency training.

The second main component of my proposed approach to nontechnical skills education in surgical residency is aimed at the continuing development of nontechnical skills throughout the course of residency training. In chapter seven, I described an approach to the formal teaching of nontechnical skills using debriefing and feedback on nontechnical performance in the OR. Nontechnical performance of senior residents, as assessed by blinded raters in the OR, improved significantly after only one session of structured debriefing and feedback from an expert facilitator. Moreover, residents perceived this type of teaching as useful and beneficial, as demonstrated in a post-training questionnaire. As a result, the Null hypothesis stated under (III) in chapter 2 was refuted, and the alternative hypothesis (IIIa) was confirmed.
8.2 Structuring nontechnical education in surgical residency training

The move toward competency-based education is changing the way we teach residents in surgical residency training programs (Sachdeva, Pellegrini, & Johnson, 2008). One of the strengths of competency-based education lies in the focus on educational outcomes that are to be attained prior to graduation (Frank et al., 2010). Educational outcomes in competency-based education are defined as skills, abilities and attitudes and, to a lesser degree, knowledge. In a learner-centered approach to education, progression through residency training, as well as the time point of graduation, are determined by the individual development of the required skills and abilities (Frank et al., 2010). This is in contrast to traditional concepts of postgraduate education that are based on a defined training duration and rely on the assessment of knowledge at predetermined time points during residency training, and upon graduation. Definitions of core competencies, or learning objectives, as well as educational milestones, or benchmarks of performance that must be achieved to progress to the next stage of residency training, are among the characteristics of competency-based curricula. In Canada, learning outcomes have been defined in the CanMEDS roles required of a medical expert, while the ACGME in the U.S. has defined core competencies to be attained in residency training. Developmental milestones are currently being defined in the U.S. milestones project (Surgery, 2014) and in the CanMEDS 2015 framework (Frank, Snell, & Sherbino, 2014). Correspondingly, the ISCP in the U.K. provides trainees and educators of the surgical specialties and subspecialties with a detailed syllabus of learning content and training outcomes, as well as assessment tools to ensure progression towards these outcomes (ISCP, 2013c). In addition, the ISCP training syllabus contains examples of relevant knowledge, skills and behaviours pertaining to communication, teamwork, and leadership (ISCP, 2013c). In contrast, the ACGME and CanMEDS 2015 competency frameworks, training requirements and milestones describe nontechnical aspects of performance relatively vaguely and in generic terms. Although abilities pertaining to competencies such as communicator and collaborator have been included in In-Training Evaluation Report (ITER) forms, it is unclear what to do with a
resident who is thought to not communicate well with patients and relatives, or to fail to “work effectively with health care workers in critical (life threatening) situations” (p.6) (RCPSC, 2010a). This reduction of entire physician roles such as communicator or collaborator to small tasks, as seen in the ITERs, but also in the PBAs, has previously been criticized as a potential area of concern with competency-based education as it is currently being implemented (Jarvis-Selinger, Pratt, & Regehr, 2012). If we want graduating surgeons to have good nontechnical skills we should not only assess compliance with the statements in the CanMEDS framework, but rather evaluate the whole range of skills, including but not limited to situation awareness; decision making in difficult situations; team coordination in the OR; the ability to establish and maintain a shared understanding among team members; leading and directing the OR team; and behaviours and attitudes pertaining to professionalism as a surgeon.

One approach to the individualized training and assessment of nontechnical skills in the context of competency-based education is formative assessment (Marriott et al., 2011). Formative assessment has previously been described as assessment for learning, in contrast to the summative assessment of learning, and has been highlighted as the primary goal of PBAs within the ISCP (Marriott et al., 2011). Evidence-based instructional strategies to enhance resident learning through formative assessments include debriefing and feedback (Cook et al., 2013; Fanning & Gaba, 2007; Raemer et al., 2011). In the study reported in chapter seven we demonstrated the effectiveness of a single half-hour session using debriefing and feedback to enhance trainees’ nontechnical performance in the OR. Integrating regular debriefing and feedback sessions in the routine workflow in the OR may be an efficient and effective strategy to ensure the development of nontechnical skills throughout residency training.

In addition, I believe that surgical trainees should develop a thorough understanding of nontechnical skills and how they relate to and complement technical skill. Establishing that understanding in the beginning of residency will provide residents with a foundation of knowledge and skills that can be expanded and refined throughout residency. The studies described in chapters five and six have revealed large inter-individual differences in innate nontechnical skills, i.e., the skill set that residents possess when they enter the postgraduate
training program, with scores ranging from the lowest possible ratings to near perfect performance. This variability in starting skills could be addressed through a basic nontechnical skills curriculum, similar to the basic training of technical skills in the Fundamentals of Laparoscopic Skills (FLS) in General Surgery (www.flsprogram.org), with the goal of providing a basic understanding of the concepts of nontechnical skills in the OR, as well as a basic set of nontechnical skills and behaviours. The knowledge and understanding of nontechnical skills would also provide incoming residents with a common language to discuss nontechnical aspects of performance with their clinical supervisors during formal and informal debriefing sessions, and to provide feedback to peers and colleagues. Integrating a nontechnical component in the formal curriculum would also underscore the role of nontechnical skills as an important aspect of surgical performance in the OR and thereby positively influence the hidden and null curricula.

In summary, my proposed approach to structuring nontechnical skills training in competency-based surgical education comprises of two main components: a basic nontechnical skills curriculum in the first year of residency training; and formative assessments throughout residency training using debriefing and feedback to ensure skill development and remediate deficiencies if observed. This approach would ensure a basic level of fundamental nontechnical skills in all residents, and document the adequate development of these skills throughout residency, while providing opportunities to remediate deficiencies.
8.3 Implementation of a basic nontechnical skills curriculum

Some crucial steps should be considered when implementing a basic nontechnical skills curriculum as detailed in chapter six. First, the curriculum should be conducted during the residents' protected study time to ensure all eligible trainees can attend. It should further be coordinated with other educational activities in the formal curriculum to avoid collisions with lectures or technical skill courses. Based on our experience, the training should also be made mandatory. Making the basic nontechnical skills curriculum mandatory underscores the importance of nontechnical skills within the formal curriculum and, more importantly, formally legitimates attendance by junior residents. The latter reason is based on observations that I made during the administration of the curriculum to two subsequent cohorts of first year residents, who volunteered to participate in the randomized controlled trial (chapter 6). Throughout the intervention, one of the biggest challenges was to achieve regular attendance by all participants. The difficulty to accomplish high attendance did not appear to be owing to a lack of motivation or interest in the topic on behalf of the residents. On the contrary, residents were highly motivated during the course sessions and expressed interest through active contributions in the seminars and group discussions. Two reasons for nonattendance were commonly reported by residents: first, the perceived requirement to be available for duties on the ward and in the OR during protected study time, when no valid reason (i.e., a mandatory teaching session) legitimized the absence; and second, concrete situations that required the resident to assist in the OR due to a shortage of staff on that particular day. Particularly the former reason for nonattendance is cause for concern, as it represents an example of exactly those expectations and norms in an educational environment that constitute the hidden curriculum and as such influence learning on the job. The hidden curriculum should also be taken into consideration when implementing a nontechnical skills curriculum or course. As described in the first chapter, the hidden curriculum comprises expectations, attitudes and opinions that exist in the learning environment. This is supported by observations made in the course of this thesis: participants in the nontechnical skills curriculum were asked in an anonymous post-course questionnaire whether they had encountered any resistance or barriers when implementing
newly acquired knowledge and skills learned in the OR, and what the nature of these barriers was. One participant responded: “Yes, often - pressure to do checklist as fast as possible is common, impatience for closed loop communication when asked of the nurses”. Another resident responded: “yes. some aren't as willing to take the time to do introductions/appropriate safety checks, less patient-focused”. In the study on debriefing and feedback, two residents reported encountering resistance when implementing concepts discussed in the sessions. One response was: “Current staff does not allow me to do checklist”. Institutional support and endorsement by clinical supervisors and role models have previously been pointed out as key factors to improve the adoption of attitudes and behaviours taught in nontechnical skills courses (K. R. Catchpole et al., 2010; McCulloch et al., 2009). In the study reported in chapter six the nontechnical skills curriculum was endorsed by the director of the residency program during an orientation for incoming residents. Future initiatives may further benefit from an involvement of faculty and senior staff members in the planning and implementation phase.

Regarding the instructional design of a basic nontechnical curriculum, the evidence suggests including various instructional strategies and methods that have been proven effective in independent studies (Flin, Yule, et al., 2007; Salas et al., 2000). Examples of instructional strategies to teach behaviours and skills include simulation (Cook et al., 2012); debriefing and feedback (Cook et al., 2013; Fanning & Gaba, 2007); behaviour modeling using video examples of nontechnical skills and behaviour in the OR environment or in other high-risk settings such as aviation (Flin, Yule, et al., 2007; Taylor et al., 2005). Following principles of adult learning, the format of instruction should be highly interactive, allowing residents to contribute previous experiences. Residents should further be encouraged to apply new knowledge and skills on the job and thus enhance experiential learning during routine practice in the OR.

Combining multiple successful instructional strategies is one of the design characteristics of aviation CRM training (Salas et al., 2000). One motivation of combining multiple instructional methods is to account for differences in individual learning styles and preferences. Differences in resident learning styles have been identified in previous studies using Kolb’s (1984) inventory of learning styles, both in General Surgery trainees (Mammen
et al., 2007) and in Orthopaedic trainees and faculty (Richard, Deegan, & Klena, 2014). Thus, accommodating different learning styles in a basic skills curriculum appears logical and advisable. More importantly though, curriculum time is scarce and work hour restrictions have substantially reduced the time available for experiential learning in the OR (Kairys, McGuire, Crawford, & Yeo, 2008). According to one study from the U.S., for instance, residents spend only 21 percent of their entire training (based on five year and 80 hour weeks) in the OR (Chung, 2005). Thus, formal courses should be designed to make an impact, and this is less likely to be achieved if nontechnical skills education is reduced to the transmission of knowledge in a lecture.

Once implemented, the basic skills curriculum should be evaluated on a regular basis to ensure processes are effective and feasible, and outcomes meet the requirements of stakeholders (Ornstein & Hunkins, 2013). Multiple different approaches to curriculum evaluation have been reported (Ornstein & Hunkins, 2013). One example is the framework for the evaluation of training programs, as proposed by Kirkpatrick (1976) and revised by Kirkpatrick and Kirkpatrick (2006), that has been described in chapter one of this thesis.
8.4 Factors influencing the Development of Nontechnical Skills

In chapter one, I discussed various factors in an educational environment that may influence the acquisition and development of nontechnical knowledge, attitudes, skills, and behaviours. External factors include the formal and informal curriculum, as well as the hidden and null curriculum. Intrinsic factors include attitudes towards concepts of nontechnical skills, as well as personality, both of which were explored in the present work.

Attitudes about teamwork and patient safety in the OR were found to be very positive among the sample of 22 General Surgery residents who completed the Team Attitudes Questionnaire during the baseline evaluation of the randomized controlled trial reported in chapter six. Residents’ agreement with the questionnaire’s statements regarding concepts of teamwork and patient safety in the OR as measured on a five-point Likert-scale (1=strongly disagree; 2=disagree; 3=neutral; 4=agree; 5=strongly agree) were very positive with median values of 4 or 5 (i.e., agree, or strongly agree) for the vast majority of items. Specifically, residents agreed or strongly agreed with statements pertaining to team structure (median, 4.50 (interquartile range, 4.00-4.75)); leadership (5.00 (4.25-5.00)); situation monitoring (4.00 (4.00-4.75)); mutual support (4.50 (4.00-5.00)); and communication (4.00 (4.00-4.50). Thus, junior residents in the beginning of their training in this program appear to have positive attitudes towards important concepts of nontechnical skills and teamwork as they relate to patient safety in the OR. Attitudes about teamwork and safety have been used extensively in aviation and healthcare to examine the culture or climate in a particular environment with regards to these concepts (Flin et al., 2006; Helmreich & Davies, 1996). In addition to a characterization of the culture, attitudes have also been interpreted as an indicator of trainees’ willingness or readiness to change behaviour as a result of a training intervention (Kraiger et al., 1993). It should be noted, however, that our results can only be used to characterize the culture among junior trainees in General Surgery, not the culture among OR staff, senior trainees, or faculty members. Anonymous surveys of surgical faculty may provide information about the organizational culture in future initiatives.

Personality traits were examined in the cross-sectional study reported in chapter five. Among the cohort of junior surgical residents we found a characteristic personality profile comprising higher-than-average scores on the domains Extraversion and Conscientiousness.
Similar findings had previously been reported from other studies on surgical trainees (Horwitz et al., 2011; McGreevy & Wiebe, 2002). Significant correlations were found between personality traits and some aspects of nontechnical performance, confirming the presence of a positive linear relationship between personality traits and nontechnical skills.

Although the demonstrated relationship of personality and nontechnical skills may be used to predict certain aspects of nontechnical performance, these data do not allow inferences regarding trainability of nontechnical skills or the development of skills during residency. Of note, the significant positive correlation between Conscientiousness and professionalism as assessed on the OSANTS scale contributed additional validity evidence to OSANTS scores on that item. Future directions for the use of personality testing in postgraduate education will be discussed in chapter 10.
8.5 Assessment of nontechnical skills

The assessment of target skills and abilities is a fundamental necessity in competency-based education (Holmboe et al., 2010). Criteria for assessment in postgraduate education include good reliability and validity of the assessments; feasibility in the target environment or context; educational effect and benefit of the assessment; and acceptability of assessments to stakeholders, including assessors and trainees (Norcini et al., 2011). As discussed previously, competency-based education focuses on skills and abilities rather than knowledge, thus assessment must be organized primarily in observations. Holmboe et al. (2010) summarized important characteristics of effective assessment in competency-based education, asserting that assessments should be frequent, with an emphasis on formative rather than summative assessment. Only through formative assessments conducted in regular intervals throughout residency can we ensure the attainment of important competencies and, more importantly, detect and remediate deficiencies in performance prior to graduation. Following Holmboe et al. (2010), assessment in competency-based education should further be criterion-referenced, and based mainly on observations in the workplace, as resident learning takes place predominantly on the job (Holmboe et al., 2010).

Three important factors should be considered when planning the assessment of residents’ nontechnical performance through observation: the context of the observations; the rating framework or tool to be used; and the assessor or observer. Nontechnical skills of surgeons have previously been assessed through observations in simulated OR environments, mainly as part of research studies (Undre et al., 2006; Undre, Koutantji, et al., 2007). Assessment in simulation has the advantage that residents’ performance can be evaluated even in uncommon and potentially dangerous situations, such as intraoperative crises (Powers et al., 2008). OR crisis simulations are also well suited for formative assessment, as they allow for individual teaching and structured feedback in debriefing sessions (Fanning & Gaba, 2007). A further strength of simulation-based assessment is the potential standardization of the scenarios, allowing for fair and comparable assessments. The downsides of assessment in simulation though are that they are costly and time consuming and require specialized personnel and equipment that may not be available to all residency programs. Workplace-
based assessments of nontechnical skills have also been reported, though equally as part of research studies (Crossley et al., 2011; Knudson et al., 2008). Of note, various forms of workplace-based assessment have already been implemented in most countries pursuing competency-based postgraduate training, predominantly by means of assessment forms such as ITERs (Ginsburg, Eva, & Regehr, 2013) and PBAs (Marriott et al., 2011) that are completed by clinical supervisors. Differences exist between assessments, for instance regarding timing. While ITERs in Canada are completed by supervisors as summative assessments of the resident’s performance on the CanMEDS roles at the end of each rotation, PBAs in the U.K. are completed jointly by the supervisor and trainee immediately after an operation with an opportunity for formative feedback (Marriott et al., 2011).

Nontechnical skills, however, are often underrepresented in these assessments, or included only as generic competencies. In-training assessment based on designated frameworks of nontechnical skills has thus far not found its way into postgraduate education. This may be owing to a lack of suitable rating frameworks or tools to score residents’ nontechnical performance in the OR. The majority of existing frameworks of nontechnical skills are aimed at the performance of entire OR teams or sub-teams, rather than individuals. Examples include the Oxford NOTECHS (Mishra et al., 2009) and OTAS (Healey et al., 2004) frameworks. An exemption is the NOTSS system that has been developed specifically to assess individual surgeons in the OR, and was subsequently used to assess surgical trainees in a large-scale study involving 85 trainees and 100 assessors (Crossley et al., 2011). NOTSS represents the most comprehensive taxonomy of surgeons’ nontechnical skills, and provides positive and negative examples of observable behaviour of surgeons in the OR. It has nevertheless not gained widespread implementation in postgraduate training.

One reason might be the substantial experience and training that is required of assessors to achieve acceptable reliability of the ratings (Yule et al., 2009). More importantly, in my opinion, is the lack of resident-specific behaviours within the NOTSS framework. In the protected educational environment of the OR trainees are rarely required to demonstrate skills such as leadership or decision making, as supervisors are always present and compensate any deficiencies before they can become visible, making the rating of residents’ performance difficult (Crossley et al., 2011). This is not a problem in simulated environments, as these allow residents to act in the role of the leader even in difficult
situations, without a staff surgeon taking over. A further potential limitation of NOTSS preventing its use in competency-based education is the design of the rating scale as a categorical four-point scale that requires the user to rate observed performance as poor, marginal, acceptable, or good, depending on the potential effect of the performance on patient safety (Yule et al., 2008). The need to interpret observed behaviours, however, may introduce a rater bias (Lingard et al., 2006). The development of the OSANTS global rating scale was prompted by the perceived need for a reliable scale to rate residents’ nontechnical performance in the OR in the context of competency-based education. The design of OSANTS is described in detail in chapter four. Most notably, OSANTS allows for the assessment of resident performance in an OR environment on seven evidence-based nontechnical skill items that are observable in trainees. Following principles of competency-based education, the rating scale used in OSANTS was designed for criterion-referenced rating, with descriptive anchors to exemplify the lowest, middle, and highest level of performance. The definitions of OSANTS items and the descriptive anchors are based on empirical evidence and describe only the most pertinent aspects of a skill to facilitate rating by less experienced users. The psychometric properties of OSANTS reported in chapter four confirmed its suitability for the in-training assessment of surgical residents in the OR, both in a simulated environment and in the real workplace.

A further important factor that must be considered in assessment is the designated rater. Crossley et al. (2011) reported using a mix of anaesthesiologists, OR nurses, and surgeons who had received minimal training in the use of NOTSS. Difficulties were reported though when active members of the OR team such as nurses and anaesthesiologists rated the performance of surgical trainees, as the observations were interrupted every time the observers were required to tend to their duties (Crossley et al., 2011). Psychologists and human factors experts were also previously used to rate nontechnical skills in the OR, although extensive prior training and practice is required to familiarize primarily non-clinical raters with routine workflow, roles, and responsibilities in the OR environment (K. R. Catchpole et al., 2007).

Raters from different backgrounds were used in the studies reported in this thesis. An advantage of using surgeons as raters is their inherent understanding of interactions and
processes pertaining to the workflow in the OR. Even simple activities such as the preoperative briefing, prepping and draping, or the roles of the primary surgeon and the assistant would have to be explained to a non-clinical rater in order to allow for meaningful assessments of the situation. One of the raters used in the study described in chapter seven was a research coordinator who was trained in the use of NOTSS and OSANTS. In addition to training in the use of the tools, the rater had to be instructed in multiple aspects of an operation such as roles of OR team members, stages of an operation, and functions of equipment, and spent hours observing in the OR to become familiar with the environment.

In a different approach to the assessment of residents’ nontechnical performance, clinical assessors could be trained and commissioned by regulatory authorities to conduct formal in-training assessments of residents in the workplace. Advantages of “official” assessors would lie in the objectivity of the ratings, as these would be conducted in a blinded fashion. Commissioned assessors could also be trained extensively in the use of the rating tools and calibrated across raters to achieve high reliability of the ratings. A potential disadvantage of formal assessments by a commissioned observer could arise from artificial effects on trainee performance caused by the presence of the rater. A designated observer who attends a case only for the purpose of assessing performance may influence a trainee’s performance either positively with the trainee acting to their best behaviour, or negatively through anxiety induced by the presence of the rater (Williams, Klamen, & McGaghie, 2003). Consequently, unobtrusive ratings may be better suited to obtain a realistic picture of trainee performance, specifically in the context of formative assessments with the purpose of documenting skill development, and remediating potential deficiencies (Williams et al., 2003). Formative assessments in competency-based residency training would ideally be performed by raters that are respected by the trainees in order for them to accept the feedback on their performance. In the study on debriefing and feedback in chapter seven surgeon-educators with experience in nontechnical skills conducted the debriefings, with very positive responses from trainees. Using surgeons as designated observers for formative assessments in the OR though is costly and may not be feasible in many residency programs (Phitayakorn et al., 2014). Observers commissioned by regulatory bodies to conduct formative assessments would be an option, although considering the number of observers that would be required to conduct frequent formative assessments of all surgical trainees
across the state or country seems like an almost impracticable and extremely costly enterprise, although I cannot provide data to support this statement. A feasible approach to the formative in-training assessment of nontechnical skills would be to train designated staff surgeons in the use of the assessment tool and integrate assessments in the routine workflow. Thus, surgeons would rate the nontechnical performance of residents while working alongside them in the OR and supervising them during procedures, and address positive and negative aspects of the residents’ performance in a debriefing and feedback session after the case. This form of informal teaching was judged as desirable by all participants in the debriefing and feedback study reported in chapter seven. In addition to the educational benefit from frequent formative assessments and feedback from individuals respected by residents, potential advantages lie in the feasibility of the assessments when performed by the direct clinical supervisor. Staff surgeons would be able to modify the degree of assistance while supervising a resident performing an operation, specifically with regards to nontechnical skills. For instance, a staff surgeon could facilitate the assessment of resident decision making and leadership skills by assuming a passive role during the case, requiring the resident to make decisions and lead the case and thereby exhibit assessable behaviours. The supervisor could also examine the trainee’s situation awareness by asking questions about the OR environment (e.g., time when operation began; availability of important equipment), the operation (e.g., how much blood has been lost), and the patient (e.g., relevant history; indication to operate; specific questions regarding relevant anatomy). Thus, the supervising surgeon has the unique ability to actively influence observability of important skills in order to obtain information about a trainee’s performance. This information could then be used to remediate deficiencies and reinforce good performance in individual debriefing and feedback sessions conducted after the operation. Feasibility and effectiveness of this approach to the formative assessment of residents is yet to be demonstrated, as will be discussed in chapter 10.
8.6 Summary

The goal of the present dissertation was to develop a structured approach to integrate nontechnical skills education in competency-based postgraduate surgical education, and to explore intrinsic factors that influence the development of these skills during residency training.

The lack of structured nontechnical skills education among surgical residency programs in North America, as confirmed in a needs analysis, was one of the principal motivators for the present work. In a cross-sectional study of incoming surgical residents from various specialties, we demonstrated large inter-individual differences in innate nontechnical performance, ranging from the lowest ratings on the rating tools used to almost perfect performance. These results indicate a need for nontechnical skills education early in residency to ensure a minimum level of performance among junior residents. Based on evidence in the literature and resident training requirements, a structured simulation-enhanced curriculum was designed to teach fundamental nontechnical skills to surgical residents in the first year of residency training. The effectiveness of the curriculum was demonstrated in a randomized controlled trial, showing that nontechnical performance of curriculum-trained residents was superior to conventional residency training. We evaluated feasibility and effectiveness of debriefing and feedback as an approach to the formative assessment of residents’ nontechnical skills in the OR, demonstrating significant improvements in nontechnical performance after a single intervention. Thus, my proposed approach to nontechnical skills training in competency-based education comprises two main components: initial training in a basic curriculum in the beginning of residency; and formative assessment in the OR in regular intervals throughout residency, with debriefing and feedback to reinforce good behaviours and remediate deficiencies of performance. In order to facilitate objective and structured in-training assessments of nontechnical performance, we developed a criterion-referenced rating tool of resident-specific nontechnical skills that are observable in an OR environment. Finally, intrinsic factors that may influence innate nontechnical skills and the development of skills following training were explored. Positive attitudes among General Surgery residents assessed in two consecutive years indicated wide acceptance of the concepts of nontechnical skills as they relate to patient safety. This can be interpreted as a willingness to change behaviour and thus
as an indicator for the success of targeted nontechnical skills training. Moreover, a positive correlation between residents’ personality traits and some aspects of nontechnical performance were identified in a cross-sectional study, suggesting the use of personality testing to define individual training needs.
9

LIMITATIONS

The main limitations of each individual study have been discussed in the limitations section of the respective chapter. In the present chapter I will discuss limitations that have not previously been deliberated, as well as general limitations of the present thesis.

9.1 Study context and participants

With the exception of the survey of North American program directors, all studies that are reported in the present work involved surgical residents from one single, large residency training program in Canada. This may have implications for the generalizability of the results, specifically when applying the findings to smaller programs with a different educational infrastructure. Conducting the studies at different sites would, however, have introduced potential confounding factors due to local differences in curricula and resident selection criteria, as well as unpredictable influences through hidden curricula that could in turn have compromised the reliability and validity of the results. A further limitation is that the majority of evidence this thesis contributes to the literature has been obtained in studies involving residents in General Surgery, potentially limiting generalizability of the results to other surgical specialties. Limiting the recruitment to residents in one program has the advantage of minimizing potential influences from confounding factors (as discussed above). An example for confounding factors is the ongoing trial of competency-based training and assessment in the Orthopaedic Surgery program, with half of Orthopaedic residents randomized to conventional training and half to the intervention.
It should be noted though that in the study detailed in chapter five we recruited junior trainees from all surgical specialties to investigate the relationship between personality and innate nontechnical skills at the outset of residency training. In that study we did not find a relationship between surgical specialty and innate nontechnical performance. Previous research using the NOTSS rating system similarly has not revealed an influence of surgical specialty or subspecialty on nontechnical performance in the OR, suggesting that nontechnical skills are relatively independent of the surgical context or specialty (Crossley et al., 2011).

### 9.2 Complex interventions

One potential limitation of the randomized controlled trial to evaluate effectiveness of a simulation-enhanced curriculum on residents’ nontechnical skills is the complex nature of the study intervention. Randomized trials of complex interventions may make it difficult to reproduce the study findings in different settings, and make it difficult, or even impossible, to specify the effective component of the intervention (Campbell et al., 2000). This criticism of complex interventions, however, was predominantly aimed at clinical trials to improve health, and not educational interventions in residency training (Campbell et al., 2000). The design of the intervention for the randomized trial conducted in the present work was guided by the theory of experiential learning and based on evidence in the education literature, specifically with regards to nontechnical education in the surgical specialties. Thus, the design largely followed current recommendations on the design of complex intervention (P. Craig et al., 2008).

Furthermore, it should be acknowledged that the purpose of the randomized trial conducted as part of this thesis was not to identify the most effective instructional method to teach nontechnical skills, but to evaluate the effectiveness of structured training of these skills on observed performance. All instructional methods used in this intervention were based on evidence and had individually been proven effective in previous studies. Since the trial was of a pragmatic nature, we aimed at conducting the most efficient and effective intervention possible, while accounting for differences in individual learning styles and preferences. Two
previous studies have explored preferred learning styles among General Surgery and Orthopaedic residents using Kolb’s (1984) learning style inventory (Mammen et al., 2007; Richard et al., 2014). These results from these studies suggest that for an intervention to be effective, training must either be highly individualized, or entail different instructional methods to accommodate for all learning styles. The latter has been the approach of choice in most nontechnical skills and CRM-training interventions, as such training is costly and time-consuming.

9.3 Long term effects and patient outcome

The third main limitation of the present work is the lack of long-term data regarding the development of nontechnical skills, as well as the lack of patient outcome data. Data from the randomized trial revealed that nontechnical skills did not improve during six to eight months of conventional residency training without targeted nontechnical skills education. It is unclear, however, how long skills will be retained in the intervention group, and how these skills will develop going forward in residency training. Questions remain with regards to the retention of nontechnical skills after initial training. The majority of studies involving simulation-based training and assessment of nontechnical skills reported conducting pre- and post-test assessment on the day of the intervention (Gettman et al., 2009; Koutantji et al., 2008), or in the case of distributed curricula, immediately at the end of the training period (Knudson et al., 2008; Marr et al., 2012). In one study of surgical residents, retention of nontechnical skills was assessed in a simulated scenario three months after the initial training, showing a significant improvement in self-efficacy beliefs and attitudes towards teamwork (J. T. Paige, Kozmenko, Yang, Gururaja, et al., 2009). McCulloch et al. (2009) also demonstrated a significant improvement in OR teams’ observed nontechnical performance and attitudes about teamwork and patient safety three months after a CRM-style intervention. Reliable data of skill retention after a dedicated training intervention in surgery is lacking. From decades of research in aviation CRM training, however, we know that nontechnical skills and attitudes do deteriorate with time and require refresher training in regular intervals to ensure continued effects (Helmreich &
One characteristic of competency-based training is to monitor the development of skills, for instance through measuring a resident’s performance against predefined developmental milestones. Thus, irrespective of the length of skill retention, mechanisms are needed in residency training to monitor nontechnical performance and identify and remediate deficiencies in time.

The lack of data regarding organizational impact of the training is a further potential limitation of the present work, since the ultimate goal of any intervention or program teaching nontechnical skills to surgeons and other OR personnel is to enhance patient safety in the OR. To date, however, prospective studies to demonstrate the effectiveness of nontechnical skills training in reducing complication rates are lacking. Although evaluation studies of the Veteran Health Administration’s Medical Team Training program have shown a decline in the morbidity (Young-Xu et al., 2011) and mortality (Neily et al., 2010) of surgical patients after implementation of MTT, it remains unclear whether this effect was due to the CRM principles that were taught in the intervention or the implementation of preoperative briefings that occurred simultaneously. Surprisingly, there is just as little evidence to support effect of nontechnical skills or CRM-training on safety in aviation, despite decades of mandatory CRM training involving tens of thousands of flight crewmembers (Salas, Burke, Bowers, & Wilson, 2001). In aviation, this has been attributed to the low incidence of major accidents such as crashes that represent the ultimate outcome variable of safety (Helmreich & Wilhelm, 1991). Surrogate measures such as incidents or near misses, on the other hand, are often biased by the training intervention as numbers of incident reports have been shown to go up after CRM training interventions, which is thought to be a caused by a raised awareness of potential issues and increased motivation to enhance safety as a result of the training (Helmreich & Wilhelm, 1991). Only few longitudinal studies have been conducted in commercial aviation to evaluate the organizational impact of CRM training. An example is the longitudinal study of Delta Airlines that, over the course of several years, reported a significant decrease in incidents and deviations after implementation of CRM training (Byrnes & Black, 1993). A potential problem of longitudinal studies over several years, however, is the impact of new technology that may enhance safety independent of crewmember performance.
In healthcare this problem is somewhat more complex, as patient outcome and mortality rates are influenced by a multitude of factors that are very difficult to quantify or even detect. Critical incident reports are potentially flawed by a reporting bias that can be due to fear of litigation or punishment, while retrospective analyses of adverse events are subject to a hindsight bias. Data collection would need to be prospective and conducted over a long period of time and involving large numbers of patients to be able to quantify the effect of surgeons’ nontechnical skills, among the multitude of confounding factors, on patient outcome in multiple regression analyses. Cook and West (2013) recently cautioned against focusing on patient outcome data when evaluating research in medical education. Among the reasons asserted by the authors was a dilution of training effects through multiple influences within a healthcare system, making it difficult to detect the effect of an intervention. This may be particularly true in education research focusing on surgical trainees in the OR, as they work almost exclusively under supervision of staff surgeons who compensate for any omissions or mistakes committed by trainees (Cook & West, 2013).

Another issue that was addressed by the authors were the sample sizes in education research that are usually not sufficient to detect a correlation with outcomes in a partial correlation involving multiple independent variables (Cook & West, 2013). Thus, for the time being, measurable improvements in the target skills and behaviours represent a reasonable and meaningful outcome after training interventions of surgical residents. Potential effects on patient outcome will have to be estimated based on empirical evidence for the role of nontechnical skills in patient safety, as discussed in chapter one.

9.4 Relationship between personality and nontechnical skills

In the cross-sectional study described in chapter five, we correlated residents’ personality as assessed by the NEO-FFI personality inventory with nontechnical performance in standardized crisis simulations. Although several associations were identified in bivariate correlations, only three remained significant after accounting for confounding factors. Questions remain whether additional significant relationships would have been detected had the sample size been larger than the 31 residents recruited for our study. The calculation of sample sizes for multiple regression analyses in psychological and behavioural research has
been a topic of controversy among research methodologists (Maxwell, 2000). Frequently, rules of thumb are used, such as the 10:1 ratio of subjects to predictors (Maxwell, 2000). When using Cohen’s (1992) formula to calculate the effect sizes of the partial correlations identified in the cross-sectional study reported in chapter 5, the resulting effect sizes fall in the medium range. Based on a medium effect size and an alpha of 0.05, the required sample size for two independent variables in a partial correlation would be 67 (Cohen, 1992). When assuming five or six independent variables, the estimated sample size according to Cohen (1992) would increase to 91 and 97, respectively. It is clear that such numbers cannot be obtained from one cohort of residents, even in a large residency program such as the University of Toronto. The population of eligible participants in our case was 56, of which 31 (55 percent) participated. This may have introduced a selection bias, for instance by selecting more extraverted individuals, as those low in extraversion may have avoided the interaction with the researchers. Recruitment, however, was proactive, in that study personnel approached potential participants based on availability within the schedule of the surgical Prep-Camp. Anecdotally, the majority of nonparticipants had a conflicting schedule of mandatory curricular components during the study period. Future studies may need to be conducted at multiple sites in order to achieve sample sizes that allow partial correlations using multiple independent variables.

One last issue regarding the use of NEO-FFI scores that should be discussed here is the risk of a social desirability bias. Surgical residency is a highly competitive environment and throughout medical school, trainees have learned to present themselves well in order to impress teachers, academic supervisors, and selection committees. Thus, it is possible that, even in the secure context of a research study with guaranteed confidentiality of study data, surgical residents may respond to certain items on the NEO-FFI in a way that conforms to general expectations of surgeon personality. Specifically transparent items aimed at Neuroticism (statements about often feeling sad and depressed, or being ashamed) or Conscientiousness (statements that one is extremely diligent; or conversely one’s inability to get things organized) may be vulnerable to “faking”. This potential weakness of NEO-FFI has been demonstrated by Topping and O’Gorman (1997), who analyzed NEO-FFI scores of 121 university students, who completed the test either as part of an exercise, or under the premise that they needed to make a good impression. Mean scores on Neuroticism,
Extraversion, Agreeableness, and Conscientiousness differed significantly between groups, indicating that NEO-FFI scores could be “faked good” (Topping & O'Gorman, 1997). This has implications for the use of personality testing in selection, as it is has been practiced in military aviation for decades. It must be assumed that highly competitive and highly intelligent individuals such as candidates applying for surgical residency are able to “fake good” some aspects of their NEO-FFI test scores. Sandal and coworkers (2005) compared NEO-FFI scores of a cohort of candidates applying for the astronaut program and compared the scores with those of active astronauts. Interestingly, the applicants scored significantly higher on Extraversion, Conscientiousness and Agreeableness, while scoring lower on Neuroticism (Sandal, Musson, Helmreich, & Gravdal, 2005). Based on their findings, the authors advised against using personality tests in selection (Sandal et al., 2005). Thus, selection based on personality may not only identify those individuals that possess desired attributes, but may also select a certain percentage of individuals who have little concern about being dishonest in a test and thus potentially have low integrity. In addition, individuals with a tendency to social desirability reporting have been found to perform poorly on problem solving tasks in challenging situations (Sandal et al., 2005). For these reasons, I believe that personality test scores, although potentially very useful when completed under honest premises to determine training needs, should be viewed with great caution when generated in the context of selection into residency training. It should be noted, however, that the previously cited studies did not identify a satisficing problem with NEO-FFI scores when these were generated under non-competitive conditions, such as the cross-sectional study described in chapter five. In our study, participants were guaranteed that NEO-FFI scores would be kept confidential and anonymized immediately after statistical evaluation, thus theoretically they had no reason to fake their responses.

Despite the limitations detailed in this chapter, I believe that the results of the individual studies presented in this thesis provide valuable information about innate nontechnical skills of junior residents, the development of these skills during the first year of residency, and the impact of structured training in a basic nontechnical skills curriculum on skill development. Moreover, the results of the studies further the understanding of intrinsic factors that may determine skill development, such as attitudes and personality. Some of the limitations
discussed here will be addressed in the following chapter on future directions for research in this area.

9.5 Influence of the Hidden Curriculum

As discussed in chapter one, the hidden curriculum plays an important role in the socialization of residents and as such may also have an impact on the development of nontechnical skills. Although we have shown in chapter six that residents’ attitudes about teamwork and nontechnical skills in relation to patient safety were very positive, we did not assess attitudes of other professional groups within the educational environment. In particular, attitudes of staff surgeons and senior residents may influence junior residents’ opinion and perception of nontechnical aspects of behaviour in the OR. Thus, prior to implementing new curricular components such as the basic nontechnical skills curriculum detailed in chapter six it may be beneficial to explore attitudes of staff surgeons and other role models in the educational environment in order to better understand the potential effects the hidden curriculum may have on junior residents. An attitudes survey such as the Team STEPPS Attitudes Questionnaire, or the Safety Attitudes Questionnaire, could be used for this purpose. To minimize negative effects of the hidden curriculum
10
FUTURE DIRECTIONS

10.1 Formative Assessments by Staff Surgeons

Formative assessment is a key feature of competency-based education. In chapter eight I have discussed my proposed approach to competency-based postgraduate education in surgery, comprising of a basic skills course in the beginning of residency, and subsequent formative assessments with debriefing and feedback on nontechnical performance throughout residency. In the present thesis, I demonstrated the effectiveness of debriefing and feedback to achieve measurable improvements in residents’ nontechnical performance. In the study detailed in chapter seven, however, the debriefing sessions were conducted by researchers. A potentially more feasible approach would be to train designated surgeons at each teaching hospital in the assessment of nontechnical skills, who would then conduct formative assessments of residents while working with them in the OR. Based on the assessments, staff surgeons could meet with the residents immediately after the shared cases and discuss residents’ nontechnical performance in a debriefing and feedback session. In a prospective study designed either as an interrupted time series as in the present work, or as a randomized controlled trial, the proposed training approach could be assessed regarding its feasibility and effectiveness.

10.2 Psychometric Properties of OSANTS when used by Staff Surgeons

The OSANTS rating scale was tested in the study described in chapter four, demonstrating good psychometric properties when assessing residents’ nontechnical performance both in a simulated OR environment and in the real OR. As pointed out in the previous paragraph though the assessment of residents’ performance in the OR may be more feasible if integrated the routine workflow. For this reason, psychometric properties of OSANTS should be tested when the tool is used by staff surgeons to rate residents’ performance in the
OR. In a prospective study, staff surgeons could be trained in the use of the tool and then complete the OSANTS rating form immediately after an operation performed together with a resident. Inter-rater agreement could be determined by calculating ICCs of several staff surgeons who rated the same sample of residents in different operations. By collecting multiple ratings of a sample of residents by several different surgeons, the generalizability coefficient could be calculated and used in a D study to determine the number of assessments needed of each trainee to obtain a reliable assessment of their performance (Crossley, Davies, et al., 2002).

10.3 Implications of Personality Testing for Surgical Education

We have identified significant associations between personality factors on the NEO-FFI personality inventory and nontechnical performance in a simulated OR. In view of the demonstrated inter-individual differences in innate nontechnical skills of incoming surgical residents, personality testing may be used to identify individual training needs and facilitate the development of tailored training programs. Conducting a similar study with a larger sample of surgical trainees, ideally involving multiple sites, could help to identify further personality traits that predict nontechnical performance in the OR and could thus be used to individualize training. An example for a personality domain that may predict nontechnical performance in crisis simulations is Extraversion. Since extraverted individuals are described as sociable, active, and assertive, and are thought to enjoy communicating with people, individuals low in Extraversion might have deficiencies in leading and directing due to the lack of assertiveness, or may not communicate effectively in the OR (Costa & McCrae, 1992). These interactions could not be confirmed in our cross-sectional study of junior residents reported here, likely due to the high scores in Extraversion among our sample. It is possible that residents with lower Extraversion scores would perform differently when subjected to the crisis simulations. A further interesting question regarding personality and nontechnical skills education is the trainability of skills. In aviation, Chidester and colleagues (1991) identified a certain personality cluster in pilots that
predicted poor results after CRM training. This is relevant to aviation CRM training, as a certain percentage of pilots (up to 5 percent) have been shown to reject the concepts of CRM and to be resistant to targeted training interventions (Chidester et al., 1991). A potential candidate domain on the five-factor model to predict responsiveness to training may be Openness, as individuals extremely low in Openness are thought to be conservative and tend to reject new thought and concepts, thus may be less willing to accept feedback on nontechnical performance and change behaviours (Costa & McCrae, 1992). Whether personality traits can be used to predict individual nontechnical training needs, or the optimal instructional approach to address these needs, are important questions that remain to be answered in future studies.

10.4 Inducing a Culture Change in the Educational Environment

The culture in an educational environment influences learners through mechanisms such as the hidden curriculum. To avoid unwanted learning outcomes such as role modelling from poor examples, the hidden curriculum should be examined, for instance by conducting attitudes surveys such as the SAQ or the T-TAQ. Birkmeyer and colleagues (2013) previously combined questions from the Hospital Survey on Patient Safety and the OR version of the SAQ to survey OR staff of bariatric hospitals, finding significant correlations between safety ratings and complication rates. Evaluating the culture in the educational environment would help to understand potential barriers and plan educational interventions and curricular components accordingly to optimize the learning of nontechnical skills in residency. For instance, strong resistance from staff surgeons against concepts of nontechnical skills would weaken or even prevent any educational effects of training interventions targeted at residents, as the vast majority of resident learning still occurs in the OR and the clinical environment in general. An isolated training curriculum would likely not be effective to ensure development of good nontechnical skills in this scenario. Thus, in order to ensure learning of nontechnical skills throughout residency, the culture in the educational environment must be taken into account. To achieve a culture change towards acceptance of the concepts of nontechnical skills as they relate to patient safety, it would be advisable to first seek endorsement from leaders and influential individuals in the
educational environment. These could be experienced surgeons who are respected by peers, program directors, as well as departmental chairs. Involving them in the development of initiatives and programs to improve patient safety will be crucial in achieving wide acceptance and ultimately a change in culture. Previous initiatives in aviation (Helmreich and Foushee, 2010) and healthcare (McCulloch et al., 2009) have shown that leadership “buy-in” is essential in the successful implementation of CRM training concepts. One approach could be to implement quality and safety rounds involving all teaching hospitals in a residency training program. These could be used as a forum to discuss topics pertaining to nontechnical skills, patient safety, and resident education, and to plan and implement quality and safety improvement measures and educational programs. Developing such initiatives “from the bottom up” rather than “from the top down” would potentially increase acceptance among all members in the educational environment. The ultimate goal of such an initiative would be an ongoing discussion and open dialogue across hospitals, and a culture that welcomes concepts of nontechnical skills.
References


Personality Inventory: results from a cohort-based trial. *American Journal of Surgery*, 201(6), 828-834. doi: 10.1016/j.amjsurg.2010.02.018


new communication routine results in improved clinical practice. *BMJ Quality & Safety, 20*(6), 475-482. doi: 10.1136/bmjqs.2009.032326


evidence for the pivotal role of clinical activities. *Medical Education, 41*(8), 763-770. doi: 10.1111/j.1365-2923.2007.02778.x


Appendices

Appendix 1: Overview of studies regarding target population, design, risk of bias, and comments including potential limitations.

<table>
<thead>
<tr>
<th>Author</th>
<th>Target group</th>
<th>Total n</th>
<th>Study design</th>
<th>Follow-up</th>
<th>Risk of bias</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arora et al. (2012)</td>
<td>Surgical residents</td>
<td>27</td>
<td>Pre-post, no control group</td>
<td>6 months</td>
<td>High</td>
<td>Self-assessment of participants. Selection bias (sample of 27 volunteers from a population of 48 eligible residents)</td>
</tr>
<tr>
<td>Brannick et al. (2009)</td>
<td>General surgery residents (PGY1-5)</td>
<td>32</td>
<td>Pre-post, no control group</td>
<td>6 months</td>
<td>Uncertain</td>
<td>Randomization of training video (before or after role play); use of blinded observers; selection bias (33 participants out of 40 eligible residents; only 23 completed pre- and post-test); Hawthorne effect possible for some outcome (reduction in error rate)</td>
</tr>
<tr>
<td>Chandawar kar et al. (2011)</td>
<td>General surgery residents</td>
<td>44</td>
<td>Pre-post, no control group</td>
<td>2 weeks</td>
<td>High</td>
<td>Observers not blinded to training status of participants (pre or post)</td>
</tr>
<tr>
<td>DaRosa et al. (2008)</td>
<td>General surgery residents in PGY 1 and 2</td>
<td>63</td>
<td>Single-blinded, randomized controlled trial</td>
<td>6 months</td>
<td>High</td>
<td>Training and assessment of decision making skills limited to laparoscopic cholecystectomy; selection bias (only 48 participants completed post-test)</td>
</tr>
<tr>
<td>Fernandez et al. (2010)</td>
<td>Surgical residents (PGY 1-3)</td>
<td>14</td>
<td>Pre-post, no control group</td>
<td>Up to 2 years</td>
<td>High</td>
<td>Selective reporting (only overall performance scores provided, no separate data on communication skills); observer not blinded to training level.</td>
</tr>
<tr>
<td>Gettman et al. (2008)</td>
<td>Urology residents (PGY 2-6)</td>
<td>19</td>
<td>Pre-post, no control group</td>
<td>Same day</td>
<td>High</td>
<td>Evaluation limited to self-assessment of residents perceived competency</td>
</tr>
<tr>
<td>Gettman et al. (2009)</td>
<td>Urology residents (PGY 1-6)</td>
<td>19</td>
<td>Pre-post, no control group</td>
<td>Same day</td>
<td>High</td>
<td>Observers not blinded to training or purpose of the study.</td>
</tr>
<tr>
<td>Hamilton et al.</td>
<td>General surgery residents (PGY</td>
<td>11</td>
<td>Pre-post, no control group</td>
<td>1 week</td>
<td>Low</td>
<td>Single-blinded observations; high interrater correlation of</td>
</tr>
<tr>
<td>Author</td>
<td>Target group</td>
<td>Total n</td>
<td>Study design</td>
<td>Follow-up</td>
<td>Risk of bias</td>
<td>Comments</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>---------</td>
<td>-------------------------------------</td>
<td>------------</td>
<td>--------------</td>
<td>--------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>(2012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rating tool</td>
</tr>
<tr>
<td>Jacklin et al. (2009)</td>
<td>Surgical residents</td>
<td>69</td>
<td>Pre-post, no control group</td>
<td>Same day</td>
<td>High</td>
<td>The same vignettes were used in the posttest as in the pretest (unclear, if decision making improved or participants remembered the individual vignettes)</td>
</tr>
<tr>
<td>Klaristenfeld et al. (2007)</td>
<td>General surgery residents (PGY 1-5)</td>
<td>47</td>
<td>Pre-post, no control group</td>
<td>3 months</td>
<td>High</td>
<td>Potential selection bias; only 34% attended all three sessions; average attendance 74%, although course was mandatory; self-report bias</td>
</tr>
<tr>
<td>Knudson et al. (2008)</td>
<td>Surgical residents (PGY 2 and 3)</td>
<td>18</td>
<td>Single-blinded randomized controlled trial</td>
<td>Not specified</td>
<td>Low</td>
<td>Observers blinded to training groups; Structured Assessment Tool used for behavioural observations was not validated; complete data set only available for 10 participants</td>
</tr>
<tr>
<td>Koutantji et al. (2008)</td>
<td>OR teams (n=34); surgical trainees of all training levels (n=9)</td>
<td>34</td>
<td>Pre-post, no control group</td>
<td>Same day</td>
<td>High</td>
<td>Observers not blinded, observation conducted by trainers.</td>
</tr>
<tr>
<td>Larkin et al. (2010)</td>
<td>Surgical residents (PGY1 and 2)</td>
<td>42</td>
<td>Pre-post, no control group</td>
<td>1 year</td>
<td>High</td>
<td>Observers not blinded; resident self-report of time management; unexplained increase in perceived stress</td>
</tr>
<tr>
<td>Marr et al. (2012)</td>
<td>Surgical residents (all PGY levels, n=30) and Emergency medicine residents (n=14)</td>
<td>44</td>
<td>Pre-post, no control group</td>
<td>Not specified</td>
<td>High</td>
<td>Observers not blinded to training or purpose of the study; uncontrolled confounding factors; observation of surrogate measures of teamwork during real-life events.</td>
</tr>
<tr>
<td>Moulton et al. (2009)</td>
<td>General surgery residents (PGY1, n=16) and Medical students (fourth year, n=16)</td>
<td>32</td>
<td>Single-blinded randomized controlled study</td>
<td>Same day</td>
<td>Low</td>
<td>Observers blinded to level of training and group assignment; participants stratified to level of training before randomization</td>
</tr>
<tr>
<td>Author</td>
<td>Target group</td>
<td>Total n</td>
<td>Study design</td>
<td>Follow-up</td>
<td>Risk of bias</td>
<td>Comments</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------</td>
<td>-------------------------------------</td>
<td>--------------------</td>
<td>--------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Paige et al. (2009 a, b)</td>
<td>OR teams (n=45); surgical residents (n=11)</td>
<td>45</td>
<td>Pre-post, no control group</td>
<td>Up to 3 months</td>
<td>High</td>
<td>Self-assessment of participants.</td>
</tr>
<tr>
<td>Peckler et al. (2012)</td>
<td>Surgical residents (PGY1, n=31) Emergency medicine residents (PGY1, n=10)</td>
<td>41</td>
<td>Pre-post, no control group</td>
<td>Same day</td>
<td>Uncertain</td>
<td>Inconsistent results of groups despite identical intervention. Selective reporting: no comparison of group baselines; possible confounding factors</td>
</tr>
<tr>
<td>Pliego et al. (2008)</td>
<td>Ob/Gyn residents</td>
<td>14</td>
<td>Pre-post, no control group</td>
<td>Same day</td>
<td>High</td>
<td>Self-report bias.</td>
</tr>
<tr>
<td>Razack et al. (2007)</td>
<td>General surgery residents (n=18) Gynecology residents (n=20)</td>
<td>38</td>
<td>Retrospective pre-post</td>
<td>Not specified</td>
<td>High</td>
<td>Self-report bias. No numerical outcome reported.</td>
</tr>
<tr>
<td>Scott et al. 2008</td>
<td>General surgery residents</td>
<td>12</td>
<td>Pre-post, no control group</td>
<td>6 months</td>
<td>Uncertain</td>
<td>Small sample size; selection bias (only 12 out of 22 participated in retention testing)</td>
</tr>
<tr>
<td>Webb et al. (2009)</td>
<td>General surgery residents (PGY 1 and 2, n=6) Non-general surgery residents (n=2)</td>
<td>8</td>
<td>Pre-post, PGY3 comparison group (n=6)</td>
<td>1 year</td>
<td>High</td>
<td>Selective reporting (no data from communication checklist provided); observers not blinded to training level or group (study or comparison)</td>
</tr>
<tr>
<td>Wetzel et al. (2011)</td>
<td>General surgery residents</td>
<td>16</td>
<td>Single-blinded randomized controlled pre-post study</td>
<td>Same day</td>
<td>Low</td>
<td>Objective physiologic parameters (e.g., salivary cortisol, coefficient of heart rate variability); blinded observers, validated teamwork scale</td>
</tr>
</tbody>
</table>
Appendix 2: Summary of study data pertaining to instructional design, learning objectives/ content, methods used for assessment, and outcome.

<table>
<thead>
<tr>
<th>Author (Reference)</th>
<th>Instructional methods</th>
<th>Course format / content</th>
<th>Assessment of knowledge, skills and attitudes</th>
<th>Relevant outcome</th>
</tr>
</thead>
</table>
| Arora et al. (2012) | Lectures | Half-day course/ Safety awareness, safety analysis, and safety improvement skills | 1) Multiple choice questions: patient safety knowledge  
2) Questionnaire: attitudes to patient safety  
3) Qualitative analysis of trainees’ observations of patient safety incidents before, and during 6 months after training | 1) Significant improvement in safety knowledge test scores from pretest to posttest  
2) Significant improvements in trainees’ perceived knowledge of patient safety, attitudes towards error analysis, and perceived ability to influence safety  
3) Enhanced awareness of patient safety issues after the course |
| Brannick et al. (2009) | Behavior modeling (Training video with short lecture and examples of good and poor behavior)  
Role-play (E.g., attention to detail in a scenario involving a patient and a nurse in the recovery room) | Single intervention/ Identification and avoidance of “human-factor” errors, with focus on situational judgment, attention to detail, and problem understanding | 1) Situational judgment test developed by the authors before training and one months after it  
2) Evaluation of recorded role-plays by blinded observers using a checklist and weighting scheme developed by the authors  
3) Prospective recording of on-the-job errors 12 months before and 6 months after training | 1) Significant improvement in measure of attention to detail; situational judgment and problem understanding improved, but not significant  
2) No significant effect of training video on role-play performance  
3) Significant decrease in number of errors and complications over time; significant decrease in percentage of complications over time; no significant decrease of index errors (errors targeted by training) over time |
<table>
<thead>
<tr>
<th>Author (Reference)</th>
<th>Instructional methods</th>
<th>Course format/content</th>
<th>Assessment of knowledge, skills and attitudes</th>
<th>Relevant outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chanda-warkar et al. (2011)</td>
<td>Lecture (Formal instruction on communication with patients) Role-play (15 min practice, followed by 15 min formative feedback)</td>
<td>90 minute workshop/ Communication with patients; discussing a cancer diagnosis with a patient</td>
<td>Observation during one of two standardized-patient encounters 1) Case-specific communication content checklist (13 items) 2) Master Interview Rating Scale (MIRS) to assess general communication skills</td>
<td>1) Significant improvement on case-specific communication from pretest to posttest 2) No significant difference between general communication scores at pretest and posttest</td>
</tr>
<tr>
<td>DaRosa et al. (2008)</td>
<td>Homework readings Group exercises (peer coaching; brainstorming session; mock malpractice trial; trigger-tape video reviews of LC) Practice with feedback (LC procedure on porcine models)</td>
<td>Three 90-minute modules during a six-month period; RCT with intervention and control group/ Intraoperative decision making during LC; human factor errors, error rescue strategies associated with critical decisions</td>
<td>Assessment of intervention and control group: 1) Knowledge test regarding critical intraoperative decisions, associated complications, error avoidance and rescue strategies (pre- and posttest) 2) Structured observation using checklist during LC in a porcine model (posttest only)</td>
<td>1) No difference between study and control group at pretest; intervention group had significantly higher scores at posttest than control; mean difference between pre- and posttest significantly greater in intervention group 2) No differences between groups to do LC on porcine model</td>
</tr>
<tr>
<td>Author (Reference)</td>
<td>Instructional methods</td>
<td>Course format / content</td>
<td>Assessment of knowledge, skills and attitudes</td>
<td>Relevant outcome</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------</td>
<td>-------------------------</td>
<td>---------------------------------------------</td>
<td>------------------</td>
</tr>
</tbody>
</table>
| Fernández et al. (2010) | Pre-simulation briefing (5-10 min, educational content regarding case to be encountered)  
Simulation (30-40 min, high-fidelity patient simulation with intra-operative crisis scenarios)  
Debriefing (10-15 min session, including video review, self-reflection, feedback, didactic teaching) | Mandatory curriculum; weekly 1-hour sessions/ACGME Core competencies: Medical knowledge; patient care; diagnosis; management; communication; professionalism | Multi-item evaluation checklists of ACGME core competencies were completed during the simulations | One year follow-up (n=10):  
Significant improvement in mean performance on checklist from PGY1 to PGY2  
Two-year follow-up (n=4):  
Significant improvement in mean performance on checklist from PGY1 to PGY3 |
| Gettman et al. (2008) | Simulation (simulated scenario with cardiac arrest; communication with relatives of patient)  
Role-play (talking to relative after unexpected patient death)  
Debriefing (feedback on simulation; group discussion) | Single intervention/Communicating difficult news | Questionnaire regarding perceived competence in communicating bad news completed by trainees | Perceived competency in communication task increased from 73.7% before to 94.7% after the simulation. |
<table>
<thead>
<tr>
<th>Author (Reference)</th>
<th>Instructional methods</th>
<th>Course format / content</th>
<th>Assessment of knowledge, skills and attitudes</th>
<th>Relevant outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gettman et al. (2009)</td>
<td>Simulation (high-fidelity simulation using intra-operative crisis scenarios) Debriefing (structured debriefing with self- and expert assessment of team performance; teamwork training videos; discussion; didactic teaching)</td>
<td>Single intervention/ CRM-principles (not specified); team performance</td>
<td>1) Expert- and resident self-assessment using Mayo High Performance Teamwork Scale 2) Faculty assessment of videotapes of simulations using a modified NOTECHS scale</td>
<td>1) Significant improvement on teamwork scale from first to second scenario in resident self-assessment and expert assessment 2) Improvement in all 19 items of the NOTECHS scale, significant in two: 1) “Adheres to best practice during procedure”, and 2) “Positive rapport maintained among all team members”</td>
</tr>
<tr>
<td>Hamilton et al. (2012)</td>
<td>Week 1: Simulation (2h session of different high-fidelity trauma simulations) Week 2: Debriefing (90mins video examples of effective/ineffective teamwork in simulations, including own simulations; group discussion) Week 3: Simulation (2h session of different high-fidelity trauma simulations)</td>
<td>Three-week course with one 2h-session each week/ Effective teamwork in trauma resuscitations</td>
<td>1) Rating of observed behaviors in week 1 and week 3 simulations using 8-item Team Function Metric 2) Survey: perceived competence as team leader in trauma resuscitation</td>
<td>1) Significant overall improvement in team function score between week 1 and week 3 2) Participants felt more competent as team leader after the course (difference not significant)</td>
</tr>
<tr>
<td>Author (Reference)</td>
<td>Instructional methods</td>
<td>Course format / content</td>
<td>Assessment of knowledge, skills and attitudes</td>
<td>Relevant outcome</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------</td>
<td>-------------------------</td>
<td>-----------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Jacklin et al. (2009)</td>
<td>Practice (assessment of postoperative risk of surgical patients in case vignettes) Feedback (cognitive feedback on participants’ assessment provided by experts; didactic content)</td>
<td>Single intervention/ Decision making; accurate surgical risk assessment; judgment</td>
<td>1) Correlation of individual judgments to gold standard as measure of accuracy 2) Mean signed error between individual judgments and gold standard 3) Mean absolute error between individual judgments and gold standard 4) Consistency measure representing participants’ internal reliability in judgment.</td>
<td>1) High degree of correlation with gold standards for at pretest; no significant improvement at posttest 2) Mean signed error at pretest 9.68%; no significant improvement at posttest 3) Mean absolute error 15.34% at pretest; no significant improvement at posttest 4) Significant improvement in consistency of trainees’ judgments from pre- to posttest</td>
</tr>
<tr>
<td>Klaristenfeld et al. (2007)</td>
<td>Lectures (between 20 and 40 min) Group discussions (between 10 and 20 min) Role-play (communicating with terminally ill patients and their relatives; feedback from staff and peers; duration 20-30 min) Reading assignments</td>
<td>3 mandatory, once-weekly 60 min sessions in 3 consecutive weeks/ Knowledge about palliative care; surgeon-patient interactions in difficult situations; breaking bad news; decision making in palliative surgery</td>
<td>15-item questionnaire (attitudes towards, and self-efficacy in questions of palliative care) completed at pretest, posttest and after 3 months</td>
<td>Significant improvement in self-efficacy in two out of 15 items: 1) “Resident received adequate training in residency about palliative and end-of-life care”: 9% agreed at pretest, 58% at posttest, and 74% at 3-month follow-up. 2) “Resident felt comfortable discussing palliative and end-of-life issues with patients and their families”: 59% agreed at pretest, 81% at posttest, and 84% at 3-month follow-up</td>
</tr>
<tr>
<td>Author (Reference)</td>
<td>Instructional methods</td>
<td>Course format / content</td>
<td>Assessment of knowledge, skills and attitudes</td>
<td>Relevant outcome</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------</td>
<td>-------------------------</td>
<td>---------------------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Knudson et al. (2008)</td>
<td>1) Lecture group: Didactic teaching (lectures and video examples of good and poor trauma resuscitations; group discussion; question and answer session) 2) Simulation group: Simulation (high-fidelity trauma simulations with different scenarios) Debriefing (Group debriefing of simulations; video examples of good and poor trauma resuscitations; question and answer session)</td>
<td>Five 2-hour sessions over 5-week period; RCT with 2 treatment groups (lecture-based and simulation-based)/ Crisis management skills in trauma situations; preparation for role as trauma team leader</td>
<td>1) Written post-training learning objectives test 2) Prospective recording of 4 real-life trauma resuscitations attended by participants as trauma team leader; assessment of recordings by two blinded observers using a structured assessment tool developed by the authors</td>
<td>1) No significant difference in learning objectives test between lecture group and simulation group 2) Simulation group performed consistently better in crisis management skills than lecture group, significantly so on the teamwork subscale; no significant difference between groups in treatment skills.</td>
</tr>
<tr>
<td>Author (Reference)</td>
<td>Instructional methods</td>
<td>Course format / content</td>
<td>Assessment of knowledge, skills and attitudes</td>
<td>Relevant outcome</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------</td>
<td>------------------------</td>
<td>-----------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Koutantji et al. (2008)</td>
<td>Interactive seminar (30 min presentation; 15 min discussion on human error, safety, role of briefings) Behavior modeling (Video examples of poor and good preparation and communication in the OR; group discussion) Role play exercises (1.5-2h classroom role-play of briefing and cross-checking; feedback from trainers) Simulation (High-fidelity intraoperative crisis simulations at baseline and posttest)</td>
<td>Single, 4.5h training session/ Human error and safety; briefings; checklists; teamwork; team communication</td>
<td>1) 14-item Briefing Attitudes Questionnaire (short version) 2) 21-item Human Factors Rating Scale, modification for surgeons 3) Modified Objective Structured Assessment of Technical Skills (OSATS) global rating scale</td>
<td>1) Shift to positive attitudes towards briefing, significant in two questions regarding the “perceived value of briefings in improving quality of care” and “perceived extent to which briefings can enhance teamwork in the operating room” 2) No significant overall effect of training on observed non-technical skills; surgeons improved significantly on decision making subscale 3) Significant improvement in surgeons’ technical skills</td>
</tr>
<tr>
<td>Author (Reference)</td>
<td>Instructional methods</td>
<td>Course format / content</td>
<td>Assessment of knowledge, skills and attitudes</td>
<td>Relevant outcome</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------</td>
<td>-------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Larkin et al. (2010)</td>
<td>Interactive presentations</td>
<td>Five 3-hour workshops distributed over the academic year/ PGY 1 curriculum: empathy and caring; time management; teamwork; informing/educating the patient PGY 2 curriculum: empathy and caring; teamwork; conflict resolution; communicating with patients with difficult personality types</td>
<td>1) Video-based test of residents’ empathy behavior (responses rated on scale developed by the authors) 2) Teamwork and Patient Safety Attitudes Questionnaire 3) Perceived Stress Scale (PGY 1 only) pre- and post-training, and at the end of the academic year 4) Self-assessment of time management skills</td>
<td>1) Significant improvement in observed empathy behavior from pretest to posttest 2) No significant changes in teamwork subscales of the questionnaire 3) Overall stress scores were significantly higher after the training 4) 67% of residents reported improvement in their time management skills</td>
</tr>
<tr>
<td>Marr et al. (2012)</td>
<td>Simulation (high-fidelity trauma simulations) Debriefing (group debriefing after each simulation session)</td>
<td>Six training sessions during 3-month period/ Trauma management; communication; teamwork</td>
<td>Pre- and post-training review of 30 video-recorded real life trauma alerts; rating of teamwork-dependent behaviors and outcomes using a checklist</td>
<td>Significant decrease in number of healthcare workers involved in trauma resuscitation Significant decrease in time to intubation from paralysis Significant decrease in time to leave trauma bay for imaging or operating room Significant increase in presence of a team leader from 64% pre-training to 90% post-training</td>
</tr>
<tr>
<td>Author (Reference)</td>
<td>Instructional methods</td>
<td>Course format / content</td>
<td>Assessment of knowledge, skills and attitudes</td>
<td>Relevant outcome</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------</td>
<td>-------------------------</td>
<td>-----------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Moulton et al. (2009)</td>
<td>Simulation (two procedural performance scenarios with standardized patients at pretest; two at posttest) - Debriefing (30 min of structured feedback from standardized patients using recorded simulations; formative feedback regarding communication and interpersonal skills)</td>
<td>Single intervention; RCT with two treatment groups (study group received debriefing after the simulation; control group no debriefing) - Communication and interpersonal skills in difficult patient interactions</td>
<td>Assessment of the videotaped patient interactions by two blinded raters regarding: 1) Communication and interpersonal skills using a validated global 5-item scale. 2) Task-specific 26-item checklist for technical skills 3) 5-item global rating scale for technical skills (5-point scale)</td>
<td>1) Between-group differences: Significantly better performance of study group than control in all subscales of communication on post-intervention scenarios. 2 and 3) No significant effects on technical performance.</td>
</tr>
<tr>
<td>Paige et al. (2009 a, b)</td>
<td>Simulation (high-fidelity simulations using intraoperative crisis scenarios) - Debriefing (reflective review of the simulations, emphasizing teamwork competencies)</td>
<td>Two modules distributed over 3 months/ - Module 1 (role clarity; open communication; shared mental models; resource management; situation awareness; anticipatory response; cross-monitoring; mental rehearsal; flattened hierarchy) - Module 2 (teamwork competencies as module 1; preoperative briefing using protocol)</td>
<td>15-item self-efficacy questionnaire targeting teamwork competencies (6-point Likert scale from 1= not confident at all to 6= completely confident), completed at pre- and posttest of both modules</td>
<td>Module 1: Significant improvement in pre/post-training mean scores for 4 out of 15 teamwork competencies Module 2: Significant improvement in pre/post-training mean scores for 9 out of 15 teamwork competencies Results were interpreted towards a cumulative positive effect of distributed training on teamwork attitudes</td>
</tr>
<tr>
<td>Author (Reference)</td>
<td>Instructional methods</td>
<td>Course format / content</td>
<td>Assessment of knowledge, skills and attitudes</td>
<td>Relevant outcome</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------</td>
<td>-------------------------</td>
<td>----------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Peckler et al. (2012)</td>
<td>Role-play (non-medical scenarios and teamwork exercises)</td>
<td>One-day workshop; two equal-sized groups of trainees completed identical workshop on different days/ Effective and ineffective teamwork behaviors; error framing and management techniques; conflict solving; crisis management; situational awareness; team leadership</td>
<td>Pre- and post-training administration of a 15-item Situational Judgment Test (behavioral responses to teamwork issues)</td>
<td>Group 1 improved from pre- to post-test, but not significantly Group 2 improved significantly from pre- to post-test with a strong effect size</td>
</tr>
<tr>
<td>Pliego et al. (2008)</td>
<td>Simulation (high-fidelity human patient simulations using four obstetric crisis scenarios)</td>
<td>“Boot-camp” during first 3 months of the academic year/ Problem solving abilities; resource utilization; interpersonal skills; leadership; communication skills; comprehension of pathophysiology; clinical competence</td>
<td>Web-based survey before and after the intervention: 1) Perceived technical competency 2) Perceived stress hardness in emergencies 3) Perceived confidence in a leadership role</td>
<td>1) Significant improvement in perceived technical competency in obstetric emergencies and neonatal resuscitation 2) Significant improvement in perceived stress hardness for all four crisis scenarios: shoulder dystocia, neonatal resuscitation, postpartum hemorrhage, ruptured ectopic pregnancy 3) Overall improvement in perceived confidence in leadership role, but not significant.</td>
</tr>
<tr>
<td>Razack et al. (2007)</td>
<td>Seminar (communication skills)</td>
<td>Single, 2.5 hour intervention/ Patient-doctor communication skills</td>
<td>4-item questionnaire of trainees’ perceived change in knowledge and skills</td>
<td>Overall significant improvement in perceived communication knowledge and skills</td>
</tr>
<tr>
<td>Author (Reference)</td>
<td>Instructional methods</td>
<td>Course format / content</td>
<td>Assessment of knowledge, skills and attitudes</td>
<td>Relevant outcome</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------</td>
<td>-------------------------</td>
<td>----------------------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Scott et al. (2008)</td>
<td>Interactive didactic lectures Practice (technical training in animal laboratory)</td>
<td>Two-day course/ Practical strategies in dealing with surgical trauma</td>
<td>Pre-training and post-training assessment of surgical decision making through a validated 45-minute case-based test Retention testing using the same test after 6 months</td>
<td>Post-training and retention-test scores were significantly higher than pre-training scores No significant difference between post-training and retention-test</td>
</tr>
<tr>
<td>Webb et al. (2009)</td>
<td>Reading assignments Interactive lectures Case-based discussions and presentations Practice (technical skills-/cadaver laboratories; interprofessional- and patient communication tasks)</td>
<td>Longitudinal curriculum for PGY 1 and 2, delivered in six 1-week blocks during protected study time/ ACGME competencies: medical knowledge; systems-based practice; patient care; professionalism; communication; practice-based learning and improvement</td>
<td>1) Multiple-choice examinations after each curriculum block regarding learning content of the respective block 2) American Board of Surgery In-training Examination (ABSITE) scores 3) Objective Structured Assessment of Technical Skill (OSATS) score 4) Presentation and communication skills observation checklist</td>
<td>1) Pre-post curriculum test scores improved significantly for PGY1 and PGY2 residents; mean results for curriculum-trained residents significantly higher than non-curriculum comparison group of PGY3 residents 2) Mean 2-year average ABSITE scores for curriculum-trained residents significantly higher than comparison group residents 3) Curriculum-trained PGY1 residents improved significantly; curriculum-trained PGY2 residents scored significantly higher than PGY3 comparison group 4) Reported improvements in communication and presentation skills between PGYs 1 and 2, but no data reported</td>
</tr>
<tr>
<td>Author (Reference)</td>
<td>Instructional methods</td>
<td>Course format / content</td>
<td>Assessment of knowledge, skills and attitudes</td>
<td>Relevant outcome</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------</td>
<td>-------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Wetzel et al. (2011)</td>
<td>Simulation (high-fidelity crisis simulation conducted at pretest and posttest) Surgical Stress Management Training (baseline information about procedure to be completed in simulation; educational video; information booklet; mental rehearsal; individual selection of coping strategies; relaxation training)</td>
<td>Single intervention; RCT with intervention and control group/ Knowledge of surgical stress management strategies; awareness of surgical stressors and coping strategies; use of coping strategies; relaxation tools for general stress management</td>
<td>Assessment of intervention and control group in pretest and posttest simulation: 1) Stress measures: State-Trait-Anxiety-Inventory (STAI) short version; coefficient of heart rate variability; salivary cortisol level; rating of observed stress; coping strategies using elements from the Surgical Coping Questionnaire; 2) Blinded observation of teamwork in simulation using the Observational Teamwork Assessment for Surgery (OTAS) scale 3) Objective Structured Assessment of Technical Skill (OSATS) score and end product assessment</td>
<td>Effects within intervention group: 1) No significant changes in STAI; tendency of lower salivary cortisol levels and observed stress in posttest simulation, but not significant; coefficient of heart rate variability significantly higher during post-intervention simulation, indicating lower mental strain after training; significant increase in coping strategies from pre- to posttest 2) Teamwork (OTAS) score of intervention group increased significantly from pre- to posttest 3) Trend to improved surgical performance at posttest in intervention group, but not significant Effects within control group: No significant changes from pre- to posttest. Between-group effects: Significant influence on number of coping strategies.</td>
</tr>
</tbody>
</table>
Appendix 3: Main outcome themes and factors considered to upgrade/ down grade strength of evidence.

### Patient-centered communication

<table>
<thead>
<tr>
<th>Author, reference</th>
<th>n</th>
<th>Quality of study</th>
<th>Risk of bias</th>
<th>Effect on communication/ Outcome</th>
<th>Factors to upgrade strength of evidence</th>
<th>Factors to downgrade strength of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chandawarkar et al. (2011)</td>
<td>44</td>
<td>Low</td>
<td>High</td>
<td>Some improvement</td>
<td>N/a</td>
<td>N/a</td>
</tr>
<tr>
<td>Gettman et al. (2008)</td>
<td>19</td>
<td>Low</td>
<td>High</td>
<td>Improvement</td>
<td>N/a</td>
<td>N/a</td>
</tr>
<tr>
<td>Klaristenfeld et al. (2007)</td>
<td>47</td>
<td>Low</td>
<td>High</td>
<td>Improvement</td>
<td>N/a</td>
<td>N/a</td>
</tr>
<tr>
<td>Larkin et al. (2010)</td>
<td>42</td>
<td>Low</td>
<td>High</td>
<td>Improvement</td>
<td>N/a</td>
<td>N/a</td>
</tr>
<tr>
<td>Moulton et al. (2009)</td>
<td>32</td>
<td>High</td>
<td>Low</td>
<td>Improvement</td>
<td>N/a</td>
<td>N/a</td>
</tr>
<tr>
<td>Razack et al. (2007)</td>
<td>38</td>
<td>Low</td>
<td>High</td>
<td>Improvement</td>
<td>N/a</td>
<td>N/a</td>
</tr>
<tr>
<td>Webb et al. (2009)</td>
<td>8</td>
<td>Low</td>
<td>High</td>
<td>Improvement</td>
<td>N/a</td>
<td>N/a</td>
</tr>
</tbody>
</table>

Strength of evidence supporting training effect on patient-centered communication was downgraded from high (1 RCT) to low (2 points due to high risk of bias in all observational studies). N= number of study participants; RCT= randomized controlled trial.

### Teamwork

<table>
<thead>
<tr>
<th>Author, reference</th>
<th>n</th>
<th>Quality of study</th>
<th>Risk of bias</th>
<th>Effect on teamwork/outcome</th>
<th>Factors to upgrade strength of evidence</th>
<th>Factors to downgrade strength of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gettman et al. (2009)</td>
<td>19</td>
<td>Low</td>
<td>High</td>
<td>Improvement</td>
<td>N/a</td>
<td>N/a</td>
</tr>
<tr>
<td>Hamilton et al. (2012)</td>
<td>11</td>
<td>Moderate</td>
<td>Low</td>
<td>Improvement</td>
<td>Blinded observers</td>
<td></td>
</tr>
</tbody>
</table>
Strength of evidence supporting training effect on teamwork was downgraded from high (2 RCT) to moderate (1 point for an inconsistency in findings and 1 point for the risk of bias in some observational studies). N= total number of study participants; RCT= randomized controlled trial.

**Surgical decision making**

<table>
<thead>
<tr>
<th>Author, reference</th>
<th>n</th>
<th>Quality of study</th>
<th>Risk of bias</th>
<th>Effect on decision making/Outcome</th>
<th>Factors to upgrade strength of evidence</th>
<th>Factors to downgrade strength of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaRosa et al. (2008)</td>
<td>48</td>
<td>Moderate</td>
<td>High</td>
<td>Improvement</td>
<td>Blinded observers</td>
<td>Selection bias (only 48 out of 63 participants completed post-test)</td>
</tr>
<tr>
<td>Jacklin et al. (2009)</td>
<td>69</td>
<td>Low</td>
<td>High</td>
<td>Some improvement</td>
<td>N/a</td>
<td>Inconsistency (no improvement of residents after training)</td>
</tr>
<tr>
<td>Scott et al. (2008)</td>
<td>12</td>
<td>Low</td>
<td>High</td>
<td>Improvement</td>
<td>N/a</td>
<td>N/a</td>
</tr>
</tbody>
</table>
Strength of evidence supporting training effect on surgical decision making was downgraded from high (1 RCT) to low (1 point for inconsistency and 1 point due to the high risk of bias in all observational studies). N= number of study participants; RCT= randomized controlled trial.

## Coping with stress

<table>
<thead>
<tr>
<th>Author</th>
<th>n</th>
<th>Quality of study</th>
<th>Risk of bias</th>
<th>Effect on stress/Outcome</th>
<th>Factors to upgrade strength of evidence</th>
<th>Factors to downgrade strength of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Larkin et al. (2010)</td>
<td>42</td>
<td>Low</td>
<td>High</td>
<td>Worsening</td>
<td>N/a</td>
<td>Unexplained inconsistency (significant worsening of perceived stress)</td>
</tr>
<tr>
<td>Pliego et al. (2008)</td>
<td>14</td>
<td>Low</td>
<td>High</td>
<td>N/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Wetzel et al. (2011)</td>
<td>16</td>
<td>High</td>
<td>Low</td>
<td>Improvement</td>
<td>N/a</td>
<td>Only effect on heart rate variability; no effects on other measures of stress.</td>
</tr>
</tbody>
</table>

Strength of evidence supporting training effect on stress and coping abilities was downgraded from high (1 RCT) to low (1 point due to serious inconsistencies and 1 point due to a high risk of bias in all observational studies). N= number of study participants. RCT= randomized controlled trial.
## Patient safety and error reduction

<table>
<thead>
<tr>
<th>Author</th>
<th>n</th>
<th>Quality of study</th>
<th>Risk of bias</th>
<th>Effect on patient safety/error/Outcome</th>
<th>Factors to upgrade strength of evidence</th>
<th>Factors to downgrade strength of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arora et al. (2012)</td>
<td>27</td>
<td>Low</td>
<td>High</td>
<td>Improvement</td>
<td>N/a</td>
<td>Selection bias (27 volunteers out of population of 48 eligible residents); self-assessment</td>
</tr>
<tr>
<td>Brannick et al. (2009)</td>
<td>32</td>
<td>Low</td>
<td>High</td>
<td>Improvement</td>
<td>Use of blinded observers</td>
<td>Selection bias (only 23 out of 32 completed pre and posttest)</td>
</tr>
</tbody>
</table>

Strength of evidence for training effect on outcome pertaining to patient safety and error reduction was downgraded from low (observational studies) to very (1 point due to high risk of bias in all observational studies). N= number of study participants.
Appendix 4: Objective Structured Assessment of Non-technical Skills (OSANTS Global Rating Scale)

**Situation awareness**

| Surgeon well prepared, monitors/makes sense of his/her environment throughout the procedure, and routinely considers future events/equipment needs. |
|---|---|---|---|
| 5 | 4 | 3 | 2 | 1 |
| Surgeon well prepared, monitors/makes sense of his/her environment, but may show occasional deterioration of situational awareness; may occasionally fail to consider future events/equipment needs. |
| Surgeon ill prepared, fails to monitor/make sense of his/her environment, completely unaware of his/her environment; repeatedly fails to consider future events/equipment needs; encounters predictable problems. |

**Decision making**

| Surgeon clearly and promptly defines a problem, generates option(s), makes a decision and implements it; reviews the outcome, if ineffective changes the plan without hesitation. |
|---|---|---|---|
| 5 | 4 | 3 | 2 | 2 |
| Surgeon defines a problem and generates option(s), but may occasionally hesitate to do so; makes/implements decisions, but occasionally appears unsure; reviews the outcome and changes the plan if necessary, but may occasionally appear hesitant/undecided. |
| Surgeon fails to define a problem, or generate option(s); fails to make/implement any decisions; fails to review the outcome, or adheres to a plan even if proven ineffective. |

**Teamwork**

| Surgeon consistently establishes and maintains shared understanding among team members throughout the operation; conducts a comprehensive briefing and surgical pause; actively encourages input/criticism from team members; volunteers to provide support/assistance if required. |
|---|---|---|---|
| 5 | 4 | 3 | 2 | 1 |
| Surgeon strives to establish/maintain a shared understanding among team members, but shows some deficiencies in the briefing/surgical pause, and/or occasional delays/failure in sharing new information; accepts input/criticism from team members, but does not actively encourage it; provides assistance/support to team members if requested. |
| Surgeon repeatedly fails to establish/maintain shared understanding among team members; omits briefing/surgical pause; fails to share new information with the team; dismisses input/criticism from team members; fails to provide support, even if requested. |
### Communication

The surgeon’s ability to ensure effective transfer of relevant information at all times by sending clear messages, articulating effectively and adjusting voice volume to ambient noise to ensure he/she is easily heard, addressing persons directly by name / role or establishing eye contact.

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Surgeon communicates effectively at all times by ensuring messages are heard and understood, sending clear and complete messages, adjusting voice volume to ambient noise, and addressing persons directly by name, or establishing eye contact.</td>
</tr>
<tr>
<td>4</td>
<td>Surgeon communicates effectively most of the time, but may occasionally send incomplete or ambiguous messages, or may occasionally fail to adjust voice volume to ambient noise and / or fail to address person directly by name, or establish eye contact resulting in occasional uncertainty regarding reception / understanding of message(s).</td>
</tr>
<tr>
<td>3</td>
<td>Surgeon fails to communicate effectively, frequently sends incomplete or ambiguous messages, fails to adjust voice volume to ambient noise, fails to address person directly by name or establish eye contact, resulting in frequent uncertainty regarding reception / understanding of messages and loss of relevant information.</td>
</tr>
</tbody>
</table>

### Leading and Directing

The surgeon’s willingness and ability to assume the role of the leader in the operating room when operating as primary surgeon (e.g., asking for own instruments) or assisting junior trainees; willingness to take charge if appropriate within a situation, and ability to use authority and assertiveness when needed.

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Surgeon consistently and clearly assumes the role of the leader while operating as the primary surgeon or assisting junior trainees, takes charge in a proactive manner when appropriate within the situation, and uses authority and assertiveness when needed.</td>
</tr>
<tr>
<td>4</td>
<td>Surgeon assumes role of the leader while operating as the primary surgeon or assisting junior trainees, but may occasionally hesitate to do so or remain passive, waiting for instructions from superior; takes charge when appropriate within the situation, but with some hesitation; may occasionally lack authority and assertiveness.</td>
</tr>
<tr>
<td>3</td>
<td>Surgeon fails to assume the role of the leader when operating as primary surgeon or assisting junior trainees, always remains passive and awaits instructions from superiors, fails to take charge even in situations when it would be appropriate, and/or completely lacks authority and assertiveness.</td>
</tr>
</tbody>
</table>

### Professionalism

The surgeon demonstrates a commitment to the patient at all times, shows accountability, is respectful towards the patient and team members, strictly adheres to standards of care and good clinical practice and through these attitudes and behaviours is a role model for team members. The surgeon maintains the aforementioned attitudes and behaviours even during stressful situations and when under pressure.

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Surgeon consistently committed to the care of the patient, accountable, always respectful towards team members and the patient, strictly adheres to standards of care, good clinical practice, and ethics and through these attitudes and behaviours is a role model for team members; maintains professional attitudes and behaviours even in stressful situation and/or under pressure.</td>
</tr>
<tr>
<td>4</td>
<td>Surgeon committed to the care of the patient, accountable, respectful towards team members and the patient, adheres to standards of care, good clinical practice and ethics, but occasionally “cuts corners”, or shows deterioration of professional attitudes and behaviours in stressful situations and/or under pressure.</td>
</tr>
<tr>
<td>3</td>
<td>Surgeon does not appear to be committed to the care of the patient, frequently shows a lack of respect for team members and the patient, disclaims responsibility for the patient; frequently cuts corners and disregards standards, behaves unethically, or shows complete deterioration or loss of previously acceptable professional attitudes and behaviours when in a stressful situation.</td>
</tr>
</tbody>
</table>
Managing and Coordinating

<table>
<thead>
<tr>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgeon organizes activities in the operating room efficiently and effectively by using all available resources (people, equipment, information, etc.) to achieve goals (e.g. by delegating tasks appropriately).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgeon organizes activities in the operating room effectively, but occasionally lacks efficiency by not using all available resources (people, equipment, information, etc.) to achieve goals (e.g. occasionally fails to delegate tasks appropriately).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgeon organizes activities in the operating room effectively, but occasionally lacks efficiency by not using all available resources (people, equipment, information, etc.) to achieve goals (e.g. occasionally fails to delegate tasks appropriately).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgeon organizes activities in the operating room effectively, but occasionally lacks efficiency by not using all available resources (people, equipment, information, etc.) to achieve goals (e.g. occasionally fails to delegate tasks appropriately).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgeon fails to organize activities on the operating room efficiently and effectively, fails to use available resources (people, equipment, information, etc.) to achieve goals (e.g. fails to delegate tasks).</td>
</tr>
</tbody>
</table>