1-1-2018

Modification of Training Methods and Alarm Thresholds: Two Ways to Reduce Potential Hazardous Clinical Alarm Related Incidents

Manikantan Shanmugham

Modification of training methods and alarm thresholds: two ways to reduce potential hazardous clinical alarm related incidents

By

Manikantan Shanmugham

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Industrial and Systems Engineering
in the Department of Industrial Engineering

Mississippi State, Mississippi

December 2018

Modification of training methods and alarm thresholds: two ways to reduce potential hazardous clinical alarm related incidents

By

Manikantan Shanmugham

Approved:

_____

Lesley Strawderman
(Major Professor)

_____

Kari Babski-Reeves
(Committee Member)

_____

Deborah Eakin
(Committee Member)

_____

Linkan Bian
(Committee Member)

_____

Stanley Bullington
(Graduate Coordinator)

_____

Jason M. Keith
Dean
Bagley College of Engineering

Name: Manikantan Shanmugham

Date of Degree: December 14, 2018

Institution: Mississippi State University

Major Field: Industrial and Systems Engineering

Major Professor: Lesley Strawderman

Title of Study: Modification of training methods and alarm thresholds: two ways to reduce potential hazardous clinical alarm related incidents

Pages in Study: 193

Candidate for Degree of Doctor of Philosophy

Within the healthcare system, nurses, are involved in many critical steps of the patient care process such as surgery triaging, post-procedure recovery monitoring and handoff release to a caregiver. A significant portion of their time is spent on the hospital floors where patients recover from their medical procedures. In today's healthcare environments, multiple devices – typically monitors, ventilators, and infusion pumps – are used during said patient recovery process. Health equipment manufacturers often add alarms to medical devices, which serve a variety of purposes, ranging from simple notifications to warnings and alerts about potential hazards that require rapid action. In typical hospital units, several types of medical devices that monitor a variety of parameters based on patient and nurses/assistants needs. Many devices have similar alarm tones, regardless of risk levels. A typical nurse will attend to multiple patients, and the number of alarms that require attention place tremendous demands on nurses' cognition, which causes enormous alarm fatigue. Alarm fatigue is not a new phenomenon and is very common in other industries, such as chemical processing, and nuclear power. The

additional stress and burden of false alarms and non-actionable alarms is also troublesome.

Many for-profit companies have developed commercial alarm management tools and aids to combat these problems and the rapid adoption of smart phones and tablets in healthcare has made alarm management more mobile and visual. However, even after these advances, the number of deaths and adverse events are still at an unacceptable level. The purpose of this study to establish that the current training methods used by various hospitals are inadequate and to explore the effects of rigorous one-on-one training and metacognitive intervention in managing alarm related adverse events. This study also identifies deficiencies in the current training methods and assesses the impact of individualizing alarm threshold settings on alarm workload, response and error rates.

DEDICATION

I would like to dedicate this research to my mother, Kannapuram and to god almighty.

ACKNOWLEDGEMENTS

A number of people have helped me to get this point. First, I would like to thank to my committee chair, Dr. Lesley Strawderman and my committee members, Dr. Kari Babski-Reeves, Dr. Deborah Eakin and Dr. Linkan Bian, for their guidance and insights throughout the project. Dr. Lesley Strawderman, as my advisor, gave me all the freedom in research to explore on my own, and at the same time supervised to recover and improve when my steps faltered. My deepest gratitude to her. They helped tremendously in formulating ideas and topics for this dissertation. Next, I would like to thank Chad Modra, Staff Vice President of Quality and Nitin Patil, Vice President of Quality, at C.R. Bard for their help and support during my coursework. Finally, I extend my gratitude to C.R.Bard for the financial support in completing all my coursework and BTG PLC for the financial support during experimentation and data collection. I would like to thank all nurses and unit aides at Swedish General Hospitals, Virginia Mason, and The Polyclinic who either participated or facilitated the data collection process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Motivation

Preventable medical errors such as medication errors, incorrect dosage, inappropriate infusions, wrong-location surgery, incorrect patient identification and missed medical device alarms contribute to 44,000 to 98,000 deaths every year, making medical errors the eighth leading cause of death in the United States (Poillon, 1999). Out of all the medical errors listed above, medical alarms are counterintuitive: the intent of these alarms is to alert healthcare providers to intervene so potential hazardous events can be thwarted; however, available evidence suggests that these alarms themselves contribute to sentinel events.

Over the past two decades, the number of medical devices in healthcare environments has grown dramatically; multiple devices in each step of the care process serve patients. Typical devices are vitals monitors, infusion pumps, ventilators, and circulatory system supporting equipment. Nurses/assistants rely on these devices to provide standard and continuous care. A study by Graham and Cvach revealed that there were 350 alarms per patient per day. This equates to 350 opportunities of committing an error per patient per day during the process of providing care. Therefore, there is definitely room for improvement (Graham & Cvach, 2010). Ever since the Joint Commission made alarm safety a National Patient Safety Goal (NPSG), academics,

nonprofits, and quasi-governmental agencies have offered numerous solutions, such as adjusting default settings, developing escalation rules, providing filters to screen-out nuisance alarms, creating tone variations, and adding middleware (for diversions to appropriate specialty nurses' units) to reduce alarm fatigue (NPSG, 2015). In addition, for-profit companies have developed commercial alarm management tools and aids. The rapid adoption of smart phones and tablets in healthcare has made alarm management more mobile and visual.

Even after these advances, the number of deaths and adverse events are still at unacceptable levels. A search, for the words "death" and "injury", in the Manufacturer and User Facility Device Experience (MAUDE) database (http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/search.cfm, last accessed Nov 28, 2017) for alarm-related issues over the course of 42 months (~ 3.5 years; April 2014 – Nov 2017) returned 189 deaths and 149 injuries. Of the 338 incidents, physiological monitors and non-life supporting ventilators contributed to 128 deaths and 90 injuries for a total of 218 adverse events – a staggering 64%. Of the 218 incidents, only 122 could be analyzed as the remaining incident reports did not contain adequate information. Of the 122 incidents, 46 (38%) were related to inadequate operator education and training and 27 (22%) were related to inappropriate work conditions. In those incidents related to operator education and training, the reports indicated that users were not completely familiar with monitoring equipment operation. This analysis clearly reveals that development of technological solutions alone will prevent adverse events.

Human intelligence is superior to machine knowledge and technological solutions in many situations. There are cases in which the technology is only as good as the people

who use it and alarm management is one such area. One possible step towards a better solution is to equip nurses and doctors with a sufficient amount of knowledge. Nurses are not always aware of the limitations of their knowledge and so cannot compensate for what they lack. Current training methods, rote memorization of critical steps and reliance on peers are no longer adequate to handle issues presented by complex devices in todays' healthcare system.

1.2 Background

Many health equipment manufacturers add alarms to medical devices, which serve a variety of purposes, ranging from simple notifications to warnings about potential hazards that require rapid action. A study on alarms by The Johns Hopkins Hospital in Baltimore, Maryland, revealed that a total of 59,000 alarm related incidents occurred at that facility over a time period of 12 days (Graham & Cvach, 2010; Cvach, 2012). In typical hospital units, several types of medical devices monitor a variety of parameters based on patient and nurses'/assistants' needs. These devices often have similar alarm tones, regardless of risk levels (Sendelback & Funk, 2013). A typical nurse will attend to multiple patients, and the number of alarms—including false and non-actionable alarms—that require attention place tremendous demands on nurses' cognition, which significantly increasing alarm fatigue.

Recognizing the importance of alarm-related deaths, the U.S Food and Drug Administration (FDA), in conjunction with the International Organization for Standards (ISO) developed alarm-related standards that address such aspects of design as human factors, usability engineering and general guidance to standardize alarms across device manufacturers (AAMI/ANSI HE 2009, IEC 62366:2015 , ISO 60601-1-8:2006). These

standards require medical device manufacturers to validate their alarms with nurses prior to commercializing.

A review of medical devices manufactured and sold post-implementation of these standards show that only 38% of devices are in full compliance and that moderate to significant differences exist among manufacturers for alarms of same device-type (Borowski et al., 2011). These discrepancies are because FDA and ISO standards cover only basic requirements. For example, the IEC 60601 specifies a maximum 10 decibels for medium-priority alarms; however, manufacturers may choose frequency and variations in pulse and tones as long as they are within the specification. A 10% below threshold blood oxygen saturation, which is a medium priority, has a constant beep tone on the Nellcor ® Pulse Oximeter with no 'burst' (quick stops) between tones, whereas the Masimo Rad-8 monitor has 'bursts' between tones. This clearly demonstrates a shortcoming of these standards. Researchers McNeer et al., (2007) argue that standardizing all properties of sound for each alarm is the only way to ensure consistency across device manufacturers. Strictly enforcing new or stricter standards across the industry would be a lengthy and extremely difficult—if not impossible—process.

Exploiting the differences in alarms, medical technology leaders such as Phillips, Amplion, Covidien, and General Electric have developed algorithms incorporated into different types of monitoring technology software (MTS). These MTSs track waveforms and numeric data and take into account a patient's clinical context, such as medications that can influence resulting waveforms or readings. The main goals of MTSs are to screen out nuisance alarms so that nurses respond only to actionable alarms and to ensure that alarms receive a timely and adequate response from caregivers. Even with this smart

technology adoption, a survey of 688 nurses found that 18% had incorrectly responded to an alarm resulting in a sentinel event in the preceding 12 months (Sendelback & Funk, 2013).

Due to 500 deaths during the preceding four-year period, the FDA and the Association for Advancement of Medical Instrumentation (AAMI) convened an Alarm Summit in 2011 with the objective to identify and develop solutions to reduce alarm-related deaths and malfunctions (AAMI Summit, 2011). Under the clarion theme (Theme # 2) of alarm system management, Dr Frank Block declared, "Clinicians do not know how the alarms work or how they are supposed to work." It is important to note that he did not identify the technology solutions or absence of industry standards as a contributor to alarm-related adverse events. Rather, Dr. Block's statement is an example of metacognitive awareness that clearly spells out an opportunity for improvement in (1) current training provided to nurses and (2) current work flow/process. His words planted the seed for this project, which was also bolstered by Solet and Barach (2012), who recommended better training for nurses to manage alarm fatigue after analyzing the phenomenon in a pediatric unit.

The Joint Commission has been addressing clinical alarm safety via National Patient Safety Goals (NPSG) and reports the collected data in its publication, Sentinel Event Alert. The April 2013 issue presented a summary of sentinel event alarm problems, and the Joint Commission presented a new NPSG for alarm management the following July (NPSG, 2013). The goal is to be implemented in two phases: Phase I (which began in January 2014) and Phase II (January 1, 2016, to year-end 2018). Phase I requires that hospitals establish alarm safety as an organizational priority and identify the most

important alarm types to manage, based on internal situations. Hospitals are also required to identify constraints and barriers in deploying an alarm management program. Even after focused efforts, as of March 30, 2017, only 90% of hospitals had completed Phase I (ECRI, 2016). The remaining 10% of hospitals have reported to the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) that the process is under way. In Phase II, hospitals are required to develop and implement specific policies and procedures to streamline alarm management and attempt to reduce or remove constraints and barriers identified in Phase I. This is likely the best time to equip hospitals with appropriate tools so that they can roll the outcome of research efforts into their Phase II process.

1.3 Metacognition and performance improvement

Though causal factors such as complex design, poor standardization, inadequate resources, poor user interface, and poor interoperability are most often cited as contributors to poor alarm system management and solutions are proposed around those contributors, one solution not previously considered for alarm system management is increasing the metacognition of nurses, particularly the aspects of metacognitive monitoring and metacognitive control. In the years following the AAMI 2011 summit, informal discussions with nurses revealed that the training provided to them is typically developed by the biomedical engineering departments of hospitals without adequate representation from nurses. Furthermore, regardless of work experience, all nurses receive the same training material: a nurse with two years of floor experience will be trained in the same way as someone who has spent a decade on the hospital floor.

It is possible that by developing an interactive one-on-one training method based on metacognitive monitoring and control, nurses will be better able to manage alarm-related adverse events, false alarms, and non-actionable alarms. Researchers in other fields have tested metacognition-based interventions and demonstrated success in test performance and timed-task performance. For example, Miller and Geraci (2011), in the field of college education, illustrated two aspects of metacognition, monitoring and control, with a simple example in their paper. A student preparing for an anatomy exam asks herself how well she remembers the bones of the hand, which is exercising her metacognitive monitoring judgment. If the information about her current state of learning is used to adjust her time spent on studying the material, then she is exercising metacognitive control. A similar relationship can be reasonably expected with respect to nurses as they are trained on alarm management. One can demonstrate his/her metacognitive monitoring ability by making a prediction on whether she he/she can recall a piece of information or large amount of information.

Joyce et al (2001) demonstrated the usefulness of metacognition in the field of high school education. After metacognition-based training intervention, study participants performed better in post-intervention tests by as much as 40%. Joyce et al.'s study concludes that teaching components and strategies of metacognition is cost-effective and provides students with a valuable skill that helps the individual become a better performer in time-sensitive tasks such as exams. Similar positive reports have been published by other researchers in fields such as chemical processing and nuclear industries (Warawun & Chokchai, 2010). Coutinho and Neuman (2008) reported that

improvement in metacognitive skills can result in good task performance. They conducted their experiment with multiple representative problems and tasks.

Despite the demonstrated success of metacognition and its usefulness in various fields, there is very little research to show whether the same level of success can be achieved in nursing care. Existing research shows that continuous training, simulations, and multi-media based training methods help reduce alarm-related sentinel events in secondary and tertiary care facilities. Although undoubtedly these training methodologies help, they are expensive and require resources such as simulators, high fidelity mannequins, seasoned trainers and money. Solet and Barach (2012) recommend better training methods for nurses involved with devices containing alarms, but they neither identify better methods nor provide information on gaps in existing methods.

1.4 Research aims

This project aimed to: (i) study the effectiveness of training methods (current) used to educate nurses on medical device alarms and (ii) assess whether providing interactive one-on-one training or feedback to nurses will influence two aspects of metacognition—monitoring judgment and control—and impact alarm management. In addition, the project also evaluated: (i) whether customization of alarm threshold limits impact response and error rates while providing care and attending various alarms and (ii) whether there is any relationship between alarm response and committed errors and perceived workload. The aforementioned objectives were achieved through three independent studies. In the first, existing training methods were replicated and compared against an interactive one-on-one training method; based on their assigned group study participants either attended lectures that were delivered in a classroom setting or took

multiple interactive one-on-one training sessions with the researcher. The classroom-training group received generic feedback and the one-on-one training group received more concrete feedback on their individual performance. Study participants' understanding of the content and subject were tested via a series of exams. In the one-on-one training session a more tailored approach was taken, and training curricula was continually adapted to nurses' needs, exam performance and competency. This study established the relationship between interactive one-on-one training and the aspects (prediction accuracy and calibration) of metacognition, monitoring judgment, and control. In the second study, the number of alarms presented to nurses were reduced, and participants' alarm response rate, committed error rate, patient care experience and overall satisfaction were assessed in the reduced alarms environment and compared against the same data collected under a default alarm environment. Non-actionable alarms were removed in the reduced alarms environment. As a follow up to the second study, participants in each alarm setting default and modified, were asked to complete a NASA-TLX sheet upon completion of their tasks. The perceived load index was calculated and its relationship with the number of alarms responded (alarm response rate), committed errors while attending alarms (error rate), patient care experience and overall satisfaction was established. The results obtained from this project could be used to develop alarm-related courses, training methods, and a work domain design program.

1.5 Research structure

The primary research question was: "Will improving nurses' metacognitive awareness help reducing alarm-related adverse clinical incidents?" To adequately answer this question and the associated research aims, three distinct studies were conducted to address the research hypotheses. The overall research structure and study layout are illustrated in Figure 1.1.

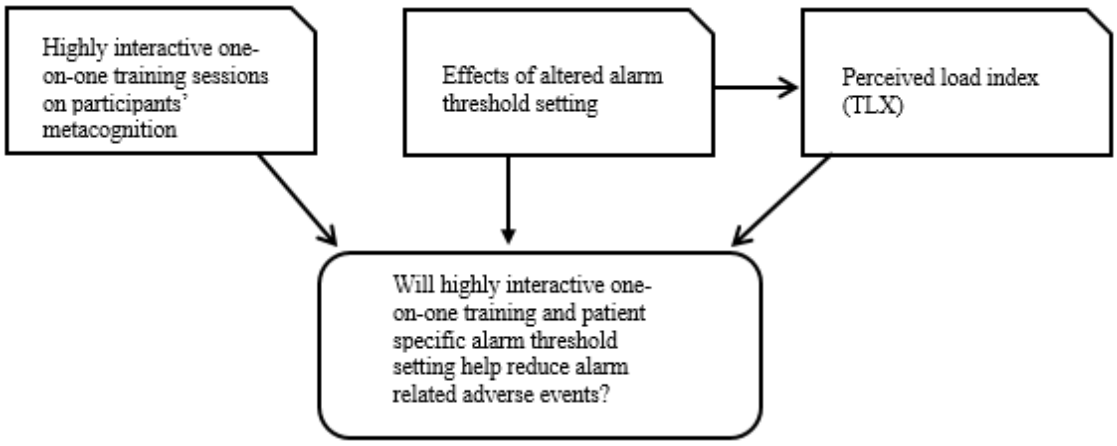

Figure 1.1 Research structure and relationship

Study 1 examined the effect of generalized and concrete individualized feedback on nurses' metacognitive monitoring, judgment and control. Study participants were randomly assigned to either classroom training or one-one-one training. Prior to attending lectures on various alarms, settings, thresholds, and scenarios, classroom participants were tested for baseline purposes. Multiple exams were administered to assess the metacognitive accuracy (any improvement) of study participants in alarm management classes. Generic feedback was given to classroom participants and they were encouraged to do better on subsequent exams. Study participants assigned to the one-on-one training group were trained in multiple, highly interactive one-on-one sessions and feedback was

individualized based on their exam performance. Similar to the classroom-training group, exams were used to assess the impact of the highly interactive one-on-one training intervention was assessed exams. Furthermore, to clearly establish the effectiveness of different training modalities, the exam performances and metacognitive prediction accuracies were compared with a non-trained study (control) group.

Study 2 examined the effects of altering alarm threshold settings on nurses' alarm response rates. A majority of the medical device manufacturers set a default value (usually textbook normal values) when they release the device to the market. Biomedical engineering staffs often do not adjust the default threshold according to patient clinical condition. The goal of this study was to examine the effect of customizing alarm threshold values based on a patient's clinical condition and on nurses' response and error rates. Furthermore, the study also assessed nurses' patient care experiences, overall satisfaction and compliance rates to procedures.

Study 3 measured the mental workload in modified alarm threshold settings and default settings and subsequently established the relationship between mental workload and alarm response rate (i.e. number of alarms responded for a total number of alarms presented) and committed error rate (i.e. number of alarms responded incorrectly). It is well known that high mental workload levels can degrade performance, and researchers have found mental workload to be a significant factor of human performance in clinical environments. Therefore, any reduction in the number of mentally demanding tasks nurses perform will have a direct impact on alarm related adverse and sentinel events. This study bolstered the fact that lower the mental workload and the fewer demanding tasks the better it is for alarm management. Potential implications from the findings of

this research are improved safety, reduced human error, and effective utilization of training resources and cost.

1.6 Real world impact

In academic settings, the concept of metacognition is a well-researched one. Many researchers have demonstrated that accurate metacognition in students is associated with better academic performance. In a landmark study, Swanson (1990) demonstrated that metacognitive monitoring abilities were directly proportional to problem-solving skills, i.e., the higher the monitoring abilities the better the problem-solving skills. Numerous studies demonstrate that individuals overestimate their knowledge or ability to execute tasks and frequently believe they are "better than average." This flawed self-assessment may lead to committing errors or result in suboptimal outcomes for tasks (Lindsey & Nagel, 2013; Hacker et al., 2008; Pennequin et al., 2010; Barenberg & Dutke, 2013; Whitebread et al., 2009; Pieschl, 2009). Such self-assessment is an aspect of metacognition. Miller and Geraci (2011), in an educational setting, demonstrated that improvements in aspects of metacognition are possible. As nurses and their aides are already highly trained and qualified professionals, similar or greater improvements in metacognition are possible in a healthcare setting.

Although improved technology, adequate staff-patient ratios, comfortable noise levels, protocol-based international standards to address alarms, and clearly defined threshold-setting all help reduce alarm fatigue and sentinel events, they are often costly and involve bureaucratic hurdles (Dunphy, et al., 2010; Fox & Riconscente, 2008; Leopold & Leutner, 2015). Metacognition-based training has not yet been tried as a solution in reducing alarm-related sentinel events. Numerous studies on alarm fatigue

reduction recommend better training as a solution because it is less expensive compared to other solutions and is generally easier to execute; however, they all fail to identify a method for improved training. This study will pave the way for developing appropriate training programs based on the needs of clinicians. Clinicians involved in managing alarms require not only high-level content knowledge but also application of that knowledge in complex situations (Tuysuzoglu & Greene, 2015; Balcikanli, 2011). For this reason, clinicians need to develop new strategies as they consider training needs for employment on hospital floors and telemetry. This research will contribute to our understanding of metacognition and offer strategies to educate nurses about medical equipment alarms with an emphasis on metacognition-based training. Any improvement, even incremental, in reducing alarm-related sentinel events will prevent harm and save lives.

1.7 References

AAMI HE 75 (2009). Human Factors Engineering—Design of medical devices. Association for the Advancement of Medical Instrumentation (ANSI). Arlington, VA.

AAMI Summit (2011). Clinical Alarms, Newsletter 2011, Arlington, VA. http://kami.camp9.org/Resources/Pictures/2011%20Alarms%20Summit%20Report.pdf, accessed November 30, 2017.

Alexander, M. J., Fabricius, V. W., Fleming, M. V., & Zwahr, M. (2003). The development of metacognitive causal explanations. *Learning and Individual Differences*, 13(3), 227-238.

Alexander, M. J., Fabricius, V.W., Fleming, M. V., & Zwahr, M., et al. (2001). The development of metacognitive causal explanation. *Learning and Individual Differences*, 13(3).

Balcikanli, C. (2011). Metacognitive Awareness Inventory for Teachers (MAIT). *Electronic Journal of Research in Educational Psychology*, 9, 1309-1332.

Barenberg, J., & Dutke, S. (2013). Metacognitive monitoring in university classes: Anticipating a graded vs. a pass-fail test affects monitoring accuracy *Metacognition and Learning*, 8, 121-143.

Borowski, M., Gorges, M., Fried, R., & Such, O., et al. (2011). Review: Medical device alarms, *Biomed Tech*, 56, 73-83.

Coutinho, S., & Neuman, G. (2008). A model of metacognition, achievement goal orientation, learning style and self-efficacy. *Learning Environments Research*, 11(2), 131-151.

Cvach, M. (2012). Monitor alarm fatigue: an integrative review. *Biomedical Instrumentation & Technology*, 46(4), 268-277.

Dunphy, B, C., Cantwell, R., Bourke, S., & Fleming, M., et al. (2010). Cognitive elements in clinical decision making. *Advances in Health Sciences Educations*, 15, 229-250.

ECRI Institute (2016). Newsletter: Compliance status of device alarm regulations, 3, 27-37.

Fox, E., & Riconscente, M. (2008) Metacognition and Self-Regulation in James, Piaget and Vygotsky, *Education Psychology Review*, 20, 373-389.

Graham, C. K., & Cvach, M. (2010). Monitor alarm fatigue: Standardizing use of physiological monitors and decreasing nuisance alarms. *American Journal of Critical Care*, 19 (1), 28-35.

Hacker, D., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection and explanatory style. *Metacognition and Learning*, 3, 101-121.

IEC 60601-1-8 (2006). Medical electrical equipment—Parts 1-8: General requirements for basic safety and essential performance—collateral standard: General requirements, test and guidance for alarm systems in medical electrical equipment and medical electrical systems. Geneva, Switzerland.

IEC 62366-1 (2015). Medical devices—Part 1: Application of usability engineering to medical devices. Geneva, Switzerland.

Leopold, C., & Leutner, D. (2015). Improving student's science text comprehension through metacognitive self-regulation when applying learning strategies. *Metacognition and Learning*, 10, 313-346.

Lindsey, B. A., & Nagel, M. (2013). Assessing students' metacognitive calibration with knowledge surveys. *AIP Conference Proceedings*, 1513, 258-261.

Manufacturer and User Facility Device Experience (MAUDE). http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/search.cfm, accessed November 28, 2017.

McNeer R. R., Bohorquez, J., Ozdamar, O., & Varon, A. J., et al. (2007). A new paradigm for the design of audible alarms that convey urgency information. *Journal of Clinical Monitoring and Computing*, 21(6), 353-63.

Miller, M, T., & Geraci, L. (2011). Training metacognition in the classroom: The influence of incentives and feedback on exam predictions. *Metacognition and Learning*, 6, 303-314.

National Patient Safety Goals (NPSG) (2013). *Joint Commission Perspectives*, 33(7), 1-4.

NPSG (2015). Hospital Accreditation Program. *The Joint Commission*, January 1, 1-17.

Pennequin, V., Sorel, O., & Mainguy, M. (2010). Metacognition, executive functions and aging: The effect of training in the use of metacognitive skills to solve mathematical word problems. *Journal of Adult Development*, 17, 168-176.

Pieschl, S. (2009). Metacognitive calibration—An extended conceptualization and potential applications. *Metacognition and Learning*, 4, 6-31.

Poillon, F. (1999). To err is human: Building a safer health system. *Institute of Medicine*, 11, 1-8.

Sendelback, S., & Funk, M. (2013). Alarm fatigue: A patient safety concern. *American Association of Critical Care Nurses*, 24(4), 386.

Solet, M. J., & Barach, R. P. (2012). Managing alarm fatigue in cardiac care. *Progress in Pediatric Cardiology*, 33(1), 85-90.

Swanson, L. H. (1990). Influence of metacognitive knowledge and aptitude on problem solving. *Journal of Educational Psychology*, 82(2), 301-309.

Tuysuzoglu, B. B., & Greene, A. J. (2015). An investigation of the role of contingent metacognitive behavior in self-regulated learning. *Metacognition and Learning*, 10, 77-98.

Warawun, S., & Chokchai, Y. (2010). Development of students' metacognitive strategies in science learning regarding nuclear energy. *AIP Conference Proceedings*, 1263 (1), 29-33.

Whitebread, D., Coltman, P., Pasternak, D., & Sangster, C., et al. (2009). The development of two observational tools for assessing metacognition and self-regulated learning in young children. *Metacognition and Learning*, 4, 63-85.

CHAPTER II

IMPACT OF GENERALIZED FEEDBACK ON METACOGNITIVE CALIBRATION: MEASURING THE EFFECT OF ONE-ON-ONE TRAINING AND INDIVIDUALIZED FEEDBACK ON METACOGNITIVE MONITORING JUDGMENT AND CONTROL

2.1 Abstract

Objective: This study aimed to assess the impact of highly interactive one-on-one training and individualized feedback on metacognitive monitoring judgment and control through medical device alarm management training classes.

Methods: Forty-five nurses and certified nurse assistants (CNAs) were randomly assigned to one of three groups: no training (control group), classroom training and one-on-one training. The control group participants did not attend any training and were asked to use their current protocol to acquire knowledge. The classroom training group attended training and simulator sessions in a classroom setting over three weeks whereas one-on-one training group covered the same material via multiple one-on-one sessions over the same three-week period. Three exams were administered during the training course. Prior to these exams, each study participant was asked to predict the score they would get on top of the answer sheet. The actual score obtained was compared against the predicted score for each participant. Impact of one-on-one training intervention was assessed via comparison of test performance between control and classroom group training sessions.

Both qualitative and quantitative data were collected and analyzed. A repeated measures ANOVA was used to quantify differences in test performance between no-training, classroom training and one-on-one training group and interaction between factors was also assessed and reviewed.

Results: Participants in all training conditions exhibited over-confident behavior at the beginning of the study (exam #1), under-confident behavior at the study's mid-point (exam #2), and at the end (exam #3). One-on-one trained participants were comparatively less over-confident at the beginning and more under-confident at later stages. Control group and classroom-trained participants exhibited no improvement in calibration by the end of the study (exam #3), whereas participants trained through the one-on-one method showed improved calibration. Results from the ANOVA showed a significant interaction between training group and exam(s). The highly interactive one-on-one training group improved their metacognitive monitoring scores over the course of 3 weeks, due to more interaction during the training intervention. Posttest reflective dialogue showed that participants used the feedback appropriately and improved their metacognition.

Conclusion: Improvement in metacognition, exam performance, and knowledge acquisition are possible through interactive one-on-one training and concrete individualized feedback. This study also shows that generalized feedback given in a classroom does not affect metacognition, and that metacognition worsened with increases in the complexity of the subject taught. Further, this study confirmed that (i) generalized feedback and suggestions provided in a classroom setting had minimal impact on study and preparation techniques and (ii) existing training methods used by hospitals do not adequately equip nurses and their assistants in managing clinical alarms.

2.2 Introduction

In a landmark study, Flavell defined metacognition as a person's "knowledge and cognition about cognitive phenomena" (Flavell, 1979). Flavell posited that monitoring and control of cognition are two functions of metacognition. One can presume that the monitoring function provides information, which is then used by the control function to monitor and select aspects of cognitive endeavor. Here is a simple example: if information about "how well one has learned a chapter for an upcoming test" is used to adjust study habits, then monitoring processes are used to control behaviors. Metacognitive monitoring and control processes allow us to observe, assess and reflect on current mental states. Individuals, who lack appropriate metacognitive skills, overestimate their performance on tests and tasks and make incorrect decisions. One explanation for pronounced overestimation in classroom settings is metacognitive deficit, such that students are generally not able to monitor their performance accurately and therefore cannot use information to alter their exam performance significantly. Not surprisingly, individuals with accurate skills prediction perform better on tasks and tests than their peers who possess lower prediction skills (Desoete, 2007; Bol & Hacker, 2001). In the past two decades, several researchers have demonstrated the usefulness of understanding metacognition in various fields. Most of these studies have been in the field of psychology and education.

Schraw described the relationship between cognition and metacognition and proposed that domain-general metacognitive practices can regulate, and, therefore, improve domain-specific cognitive tasks, such as clinical decisions, mathematical problem solving and scientific puzzles. Using Schraw's relationship, if one could

influence the outcome or result of domain specific tasks in the nursing field, the impact would be strongly positive, particularly on patient safety. Having accurate metacognitive monitoring abilities is important in nursing settings for several reasons. Most importantly, metacognitive monitoring ability directly correlates with better clinical task performance during the training period and better patient outcome later. To function effectively in a complex healthcare system, nurses must be skilled, know how to learn, and know how to apply their knowledge when situations are presented. Accurate monitoring of new learning enables nurses with effective metacognitive strategies to concentrate on domain-specific knowledge and adjust their goals and expectations. During internships, clinical rotations, or advanced targeted training programs such as medical alarms management, students must absorb a great amount of new material in a limited amount of time. Those who accurately distinguish between what they have learned and what is remaining or unknown have an advantage in providing patient care. However, unfortunately, many nurses have ineffective metacognitive strategies when it comes to learning. Therefore, it is imperative to evaluate nurses' metacognitive abilities and tailor instruction methods to the development of these core-learning strategies.

2.3 Background

Nursing care is changing dramatically given the need for nurses and their assistants to address complex clinical conditions and multiple patient comorbidities. Nurses frequently experience difficulty applying knowledge gained from didactic instruction and on-the-job training provided by their employers to make important clinical decisions for optimal patient care. To function effectively in this ever-evolving, complex healthcare system, nurses must learn to be skilled thinkers, know how to learn,

and know how to apply what they know in clinical situations. Mere memorization of formulas, facts, and outcomes from case studies and reliance on co-workers or employers to prescribe how to apply theoretical information no longer serves as adequate preparation. Teachers in the field of nursing and medicine are often encouraged to transition to new teaching and learning paradigms to address expanded needs and to keep up with technological advances. New methodologies, such as learner-centered teaching (LCT) and problem-based learning, support an active student role in learning and assist students to move from a basic understanding of information at the knowledge and comprehension levels to a higher level of understanding. Therefore, it is possible for nurses to learn more from the clinical setting with the incorporation of metacognitive practice.

Metacognitive knowledge and skills are linked to problem solving performance. The more individuals control and regulate the strategies they use, the better their capability to solve a problem (Swanson, 1990). Metacognitive processes allow people to choose strategies explicitly by thinking about their understanding of demands, and their available resources. Metacognition refers to higher-order mental processes that are often involved in learning: making plans for learning, monitoring learning speed, and predicting performance (Hacker, 1998). During the forethought, performance, and self-reflection phases of Self-Regulated Learning (SRL), metacognitive processes control learner choices and manage outcomes.

In the context of self-regulated learning, students need an adequate impression of what they have or have not learned to regulate their learning behavior effectively (Bol et al., 2005). Accordingly, they need to monitor their learning process, recognize problems,

and relate findings and observations to learning strategies. The possession of good metacognitive abilities is thought to improve performance (Thompson et al., 2011). Individuals who can accurately judge their learning are more effective learners and hence better task performers.

Few studies have shown that metacognitive knowledge and skills can be trained successfully (Huang et al., 2012; Jiang & Kleitman, 2014; Pennequin et al., 2010), and these studies were predominantly in mathematics, physics, and psychology. Researchers found that self-judging improved with training (Hacker et al., 2000). In addition, researchers have attempted to improve metacognitive accuracy with the goal of improved performance. Findings of these studies are somewhat mixed; though metacognitive accuracy or prediction accuracy improved, knowledge or content knowledge did not improve significantly (Leybina & Skvortsova, 2009). Researchers indicate that multiple sessions of interactive one-on-one training may not only improve metacognition accuracy, but also content knowledge (Ibabe & Juaregizar, 2010); however, no solid evidence exists to support this premise. This under-researched area is the foundation for the current study.

Due to the overarching nature of metacognition, it is difficult to assess quantitatively. Researchers have employed self-reports, observational methods, trace data, and monitoring judgements (Bielaczyc et al., 1995). Monitoring judgements are obtained through confidence measurements taken in real-time during a test and then transformed into confidence scores or calibration scores that indicate the match between perception of and actual level of performance. In one study, repeated practice of performance predictions improved calibration or metacognitive accuracy (Kelemen et al.,

2000). In their work, Miller and Geraci stated "when people make metacognitive monitoring judgements about particular facts on an item-by-item basis, their accuracy is measured by a correlation coefficient; this measurement is referred to as resolution. In contrast, when people make monitoring judgements about a large number of items, accuracy is measured by the degree to which the prediction corresponds to the actual level of performance; this type of measurement is referred to as calibration" (Miller and Geraci, 2011; pp. 457-458). They reported that students who performed better on exams tended to more accurately predict the score they would obtain compared to their counterparts who obtained lower scores. Their suggestion is that participants' use of feedback may vary according to the extent to which they externalize negative outcomes. In a study conducted by Cao and Nietfeld (2006), high-achieving students showed a clear advantage in self-efficacy over their low-scoring counterparts. During reflective dialogue conducted in the post-study phase, participants indicated that generalized feedback was a key contributor to regulating their behavior. Low-achieving students displayed a lack of control of their learning, suggesting that their achievement relied on individual metacognitive abilities (Conway, 2005). This indicates an opportunity for further improvement: through individualized feedback related to abilities, metacognitive skills may improve over time. Since customized feedback is possible in a one-on-one setting, improvement in metacognition will be significant for this type of learning.

2.4 Methods of teaching nursing professionals

Students in nursing, medicine, and engineering in the United States are known to be diverse in terms of ethnic background, culture, and learning style. Students in a typical nursing program include traditional and nontraditional students, male and female

students, and students of different cultures, each with a different learning ability and style. Educators, not only in nursing but also in other fields, often struggle with the challenge of teaching these students a large volume of content in a short amount of time. Designing lesson plans that accommodate the diversity of learning styles can be time intensive for educators, who are responsible for ensuring that students retain information and are able to apply knowledge learned in the classroom in the clinical setting. Students' needs, learning styles, and abilities in traditional classroom settings vary significantly and can be difficult to accommodate. Due to individual differences in knowledge, motivation, and exposure, teachers encounter difficulties in creating optimal teaching plans that cater to all students in their classes. This has motivated researchers to explore alternate teaching methods such as one-on-one teaching, individualized student instruction, or self-paced on-line instruction methods.

The traditional lecture presentation is perhaps the most well-known and often used teaching strategy regardless of differences in the student population; researchers frequently assess this method for its effectiveness and appropriateness. Students have adapted to the classroom lecture method in their learning process and have come to rely on it for gaining necessary knowledge. Further, students may have an increased comfort level with this traditional teaching methodology partly because they can remain in a passive role. Although existing research supports the use of classroom teaching as an effective teaching modality, nurse educators continue to search for more effective ways of teaching and transferring applied technology skills. It is well established that there are benefits associated with classroom teaching and lectures such as clarification of abstract concepts, organization of thinking, and development of methods of problem solving.

However, some concerns exist about the classroom-based teaching method. Information may be blindly memorized, making transfer and generalization difficult. When students encounter new problems, they are unable to adapt what they have learned, are not flexible, and tend to rely on inappropriate strategies or rote knowledge. Nursing education, in general, has traditionally focused on the lecture-based knowledge transfer model. However, the Institute of Medicine and National League for Nursing have recommended a move toward LCT and the use of different pedagogies that can enhance student learning and success (http://www.nln.org/docs/default-source/professional-development-programs/excellence-in-nursing-education-model-(pdf).pdf?sfvrsn=0, last accessed Dec 17, 2017).

2.4.1 One-on-one teaching

Education determines a healthcare professional's career and economic future and is significant to their intellectual and professional development, sense of identity, and sense of place in the clinical world. However, little research has focused on which teaching methods used in didactic settings are related to positive student achievement and better patient outcomes, and on methods of educating nurses and how they correlate to patient safety and better clinical outcomes. Studies have demonstrated the value of small group teaching and training by peers and their relationship to healthcare professional's academic and job performance (Rawson, 2011; Hamid & Mahmood, 2012). However, these studies do not focus specifically on patient safety, clinical outcomes, or overall satisfaction. Therefore, understanding the impact of modifying training methods could guide educators in developing better training methods for teaching technical content such as alarm management. In the classroom, the roles that educators and students play are less

diverse, easier to identify and categorize, and largely directed toward the acquisition and retention of course content. However, in other teaching methods such as one-on-one and peer-to-peer, interactions are often complex. Depending on the setting and goals, they may function as teaching–learning encounters, mentoring opportunities, supervisory sessions, consultative relationships, or opportunities to motivate, coach, and guide students.

2.5 Feedback mechanism

An indispensable part of an effective teaching–learning environment is providing appropriate guidance and feedback to students. Though there is no clear definition of feedback in higher education, it is universally described as an interactive process, which intends to provide learners with insight into their performance. Feedback ranges from providing grades to understand performance to offering guidance on the knowledge and skills needed for future performance (Brookhart, 2008). According to Nicol and Macfarlane-Dick, there are several direct benefits of providing proper feedback: "(1) helps clarify what good performance is, (2) facilitates the development of self-assessment in learning, (3) delivers high-quality information to students about their learning, (4) encourages teacher and dialogue around learning, (5) encourages positive motivation beliefs and self-esteem, (5) provides opportunities to close the gap between current and desired performance, and (7) provides information to teachers that can be used to help shape teaching" (Nicol and Macfarlane-Dick, 2006, p205-206). Nicol and Macfarlane-Dick suggest that student-centered rather than tutor-controlled feedback be provided. When feedback is student-centered, students are engaged in a continuous process of assessing and reflecting on their work and the entire feedback process becomes self-

regulated. Weaver (2006) demonstrated that students attached more value to the comments and dialogue provided by their lecturers than to the grades received. Few researchers recommend using feedback as a supplemental method to reinforce abstract concepts taught in traditional lectures. Studies conducted in the United Kingdom found that students are often dissatisfied with feedback, namely in terms of its accuracy, timeliness, and content (Carless et al, 2011). Ott et al indicated that large class sizes and diverse student learning backgrounds are today's main challenges in providing students with quality feedback (Ott et al, 2016). Research shows that there is no clear set of rules and guidelines for "effective feedback," and no consensus exists among educators. However, it is widely accepted that educators should acknowledge and respond to the learner's personality and provide detailed, timely feedback at the individual level. Other characteristics such as the environment (classroom vs. private), language, format, and delivery may also be of importance. It is widely acknowledged that, when feedback is contextualized, it is more likely to be accepted by students without resistance.

2.5.1 Types of feedback

There are two types of feedback provided by educators to students: constructive, which generally highlights "negative" aspects, and reinforcing, which generally revolves around "positive" aspects. Whether positive or negative, feedback should always be an unbiased reflection that logically connects with concepts and imparts knowledge (Smith, 2005). It should enable the learner to change or modify their practice and behavior and become effective practitioners. Duffy and Hardicre highlighted that if feedback is to become part of the learning process, it is essential that educators provide appropriate information that enables students to recognize clearly the strengths and weakness of their

work (Duffy & Hardicre, 2007; Lunenbrg, 2010). Positive feedback includes general praise, describing how the strengths in a learner's work match expectations, and how those strengths correlate to students' learning process. It also points out items the learner could improve upon. Negative feedback generally points out errors, criticisms, poor work, and remedial actions needed from the student (Winne et al., 2001). A mixture of positive and negative feedback types are often used by educators in various fields, especially healthcare. Educators begin with positive aspects, highlighting areas in which the student did well, and then shift to pointing out errors and conclude with suggestions for improvement (Hamid & Mahmood, 2010). Feedback on clinical practice and class performance is important for effective learning in a nursing course. Acute awareness of students' needs and understanding of the elements of the feedback process can aid the learning process and ensure that educators and nursing students have an enriched learning environment.

2.5.2 Role of feedback in nurse education

One of the primary responsibilities of nurse educators is to provide feedback to their nurses and wards that will result in meaningful patient outcomes. Feedback is vital to ensure that the student develops his or her clinical practice. To reap the benefits of clinical education, which is part of nursing curricula, feedback to students should be provided (Good Practice, 2013; Gray & Smith, 2000). Such feedback should provide the student with information on current practice and clinical task execution and offer practical advice for improved performance. Feedback given by nurse educators should provide an unbiased critique of performance, and reflect on examples and events as they occurred, with the intention to rectify mistakes and errors and increase clinical

knowledge and understanding of the subject. Educators are required to assess and provide feedback on such aspects as students' applied knowledge base, interpersonal skills, safe clinical skills, and attitude towards resolution of conflicts (Begley & white, 2003; Wood, 2000). Research shows that feedback increases students' motivation, confidence, and self-esteem.

Within the assessment process, feedback is an important and powerful part of influencing future learning (Koh, 2007; Kuiper & Pesut, 2004). Through appropriate feedback techniques, educators aim to motivate students to want to learn and develop; however, not addressing performance issues may hinder students' development and application of skills in clinical settings (Lunney, 2008). Feedback that is considered unhelpful to improve learning includes comments that are too general or vague and lack guidance.

Quality feedback thus plays a critical role in nurses' learning process in the classroom and clinical practice (Allen & Armour-Thomas, 1991). It is suggested that by considering the context of expectations and learning outcomes, and providing clear and timely guidance on actions to be taken, educators could greatly improve the value of feedback (Padden, 2013). Development of personalized learning environments is among the most important research areas of health sciences education for the next decade (Ewing, 2005; Gao & Quitadamo, 2015). Such environments should be capable of accurately tracking learners' activity, monitoring their individual characteristics, and intervening with focused feedback to improve learners' performance.

2.5.3 Tailored feedback for learning

Learners differ from each other in many ways including prior knowledge, meta-cognitive skills, motivational and affective state, and learning strategies and styles (Soderstrom & McCabe, 2011). As there are many individual factors that may influence how feedback is processed by each learner, providing high-level generic feedback or feedback lacking substance will be of less value (Ellis, 2016). The main source of information that guides a learner in their learning efforts is feedback on performance in class, through tests and assignments. Therefore, the focus of feedback provided by educators should address at least two dimensions: content mastery and tips or tools for effective learning. It is easier to give generalized feedback it is to give individualized or tailored feedback. However, research shows such feedback may not be effective as it may not be the most suitable for individual students and not necessarily address the root cause of issues (Govaerts, 2008; Jeffries, 2012; Yaeger & Arafeh, 2008). Two key benefits of tailored feedback are that it can be specific to a student's situation and private. Further, research shows that no association exists between the quality of negative feedback and self-monitoring and that nursing students are willing to accept constructive and negative feedback when it is specific to their situation (Farrell et al, 2015). In other words, information provided by the educator will be of maximum usefulness and the student will not have to worry about peers' reactions or opinions.

Seeking feedback is a behavior closely associated with self-regulated learning; a few researchers have demonstrated that students who seek feedback tend to perform better on assignments and possess metacognition skills. However, for students to seek feedback proactively, the learning environment has to be conducive. Research by

Gutierrez and Schraw (2015) shows that students proactively seek feedback and engage in discussions in tutoring, groups, or smaller class settings more than they do in large groups or traditional classroom settings. There are several external and internal processes that affect the willingness of the student to take responsibility for becoming proactive in their learning, including intrinsic motivation (Fawcett, 2014). Due to their diverse educational backgrounds and different training methods, nurses' application of knowledge tends to vary. Therefore, one type of feedback for the entire learner pool may not yield fruitful results and has the potential to result in negative consequences such as demotivation and poor self-regulated learning (Clarke, 2012). Given that they are the focal point of the learning activity, learners should have a central role in determining the feedback content and process. In other words, feedback must be learner-centric and should be based on the individual learner situation.

Plakht et al. (2012) concluded that "high-quality positive feedback, as rated by nursing students, is associated with higher achievements, higher contribution of the clinical practice to the student and over-self-evaluation. Whereas high-quality negative feedback is related to an accurate self-evaluation of the students' performance." They further recommended providing appropriate amounts of both positive and negative feedback, as well as asking educators to close the gap between current (observed) behavior and desired behavior, through a conducive atmosphere, which can be inferred to mean private and smaller settings. They warn that feedback provided should not create room or opportunity for over-estimation of one's performance and encourage constructively tailored, focused, and negative feedback, most easily accomplished in a one-on-one setting and private learning atmosphere.

2.6 Metacognition and educating nurses

The work environment for nurses often requires not only high-level content knowledge, but also application of that knowledge in complex situations with minimal or no help. For this reason, nurses need to develop new strategies as they think about their learning (McAllister et al., 2013). Research in the field of metacognition may offer a useful framework to improve nurses' learning process. Metacognitive strategies are "higher order executive skills that may entail planning for, monitoring, or evaluating the success of a learning activity" (O'Malley and Chamot, 1990: p44). O'Malley and Chamot (1990), for instance, have differentiated the range of cognitive strategies into two main types: metacognitive and cognitive. Metacognitive strategies oversee, direct, and regulate the learning process. Cognitive strategies refer to approaches students use to process new information from texts and lectures into short- and long-term memory (Greene & Azevedo, 2010 Thiede et al., 2012; Winne & Nesbit, 2009). When processing information within the classroom-learning environment as well as during independent study, students use such strategies. Self-regulated students control their learning experiences through the use of a variety of strategies that are identified as metacognitive or cognitive (Duncan & McKeachie, 2005).

Anderson (2003) demonstrated that metacognitive strategies play a more significant role than other learning strategies in the learning process, because once a learner understands how to regulate his/her own learning through the use of strategies, the learning process generally accelerates and the learner acquires and retains more knowledge. Strategic learners have metacognitive knowledge about their own thinking and learning approaches, a good understanding of what a task entails, and the ability to

orchestrate the strategies that best meet both the task demands and their own learning strengths (de Bruin & van Gog, 2012).

Developing metacognition brings learners an awareness of the learning process and strategies that lead to success. When learners are equipped with this knowledge, they understand their own thinking and learning process, and, accordingly, are more likely to oversee the choice and application of learning strategies, plan how to proceed with a learning task, monitor their own performance on an ongoing basis, find solutions to problems encountered, and evaluate themselves upon task completion (Maki et al, 1990; Maki et al, 2005). Rahimia and Katal indicated that "metacognitive knowledge is crucial for learners selecting and activating strategies and it is important that teachers strive to develop students' own metacognition and teach them how to use strategies that they find effective for the kinds of tasks they need to accomplish in the process of language learning" based on their work in assessing metacognition (Rahimia and Katal, 2010). This is similar to the discovery of Sart (2014) in his work on the development of metacognition. Sart's research reiterates that metacognition "is mindful engagement of the user in a task, including the knowledge and control the user has over his cognitive processes." It deals with awareness, observation, reflection, and analysis, which is needed to become an independent learner.

Basic metacognitive strategies include connecting new information to old information; selecting deliberate thinking strategies; and planning, monitoring, and evaluating thinking processes (Chua et al, 2012; Melby-Lervag & Hulme, 2013). They help learners regulate and oversee learning activities such as taking conscious control of learning, planning and selecting strategies, monitoring the process of learning, correcting

errors, analyzing the effectiveness of learning strategies, and changing learning behaviors and strategies when necessary.

Kuiper investigated metacognition as it is applied in nursing and nursing education (Kuiper & Pesut, 2004). Their work shows that self-regulated learning strategies improves cognitive and metacognitive skills in clinical contexts through effective clinical reasoning and reflection. Students, in their study, made significant gains in self-observation, self-judgment, knowledge work and use of health care personnel resources through 'contextual learning' model. Metacognition interventions in the field of education, nuclear science, and psychology have been shown to improve learning and cognitive processes, which could possibly be generalized to other domains and learners (Rawson et al., 2011; Serra & DeMarree, 2016). Therefore, the explicit incorporation of metacognition training for practicing nurses may be an effective strategy for promoting learning in continued nursing education that can directly result in improved critical thinking and better patient outcomes.

Nurses entering hospitals and clinics are products of a lecture-driven education system in which memorization and regurgitation of information in a scrutinized environment are generally considered indicators of success. Nurses spend a significant amount of time during early rotation years switching roles in classroom training. Nurse educators should capitalize on this class time by making nurses responsible for their own learning and allowing them to take ownership of their learning processes (Thomas & Walsh, 2008). However, nursing research on metacognition and education is mostly related to general clinical practice and decision-making. This study is the first to analyze the impact of a metacognitive training intervention for medical device alarms.

This research will contribute to our understanding of metacognition-based training and help nurse educators and biomedical trainers in developing appropriate training methods and instructional strategies in alarm management courses. Subsequent studies with this cohort of participants could explore the transfer of metacognitive skills to real time applications and pedagogical techniques necessary to promote metacognition in nursing school.

2.7 Assessment of metacognition

Metacognitive skills can be viewed as the voluntary control people have over their own cognitive processes. Zabrucky et al. (2009) found that metacognition is comprised of metacognitive knowledge and metacognitive experiences, both of which are important to learning and performance. Metacognitive experiences involve, in part, students' awareness of progress of cognitive tasks. They not only help learners in progress monitoring but also aid them in using appropriate strategies to achieve progress and alter study habits and behavior (Szpunar et al., 2014; Desoete & Roeyers, 2006). To assess these metacognitive components, that is, knowledge about strategies, and their relevance, usage, and application in certain situations, researchers in this field have used various methods (Coutinho, 2007). Researchers have done considerable work in finding how people monitor their progress during the learning process, with the hope that products of metacognitive monitoring guide learners' decisions and they choose appropriate strategies in acquiring knowledge.

2.7.1 Calibration

According to Pieschl, calibration “is a metacognitive monitoring skill and often refers to the accuracy of learner's perceptions of their own performance" (Pieschl, 2009, pp.4-5). Several researchers have demonstrated that accurate metacognitive monitoring is a prerequisite for successful learning outcomes (Alkan & Erdem, 2012; Batang, 2015; Memnun & Hart, 2012; Mokhtari & Reichard, 2002). Calibration, as generally defined in the literature, refers to the relationship between a learner's level of confidence in their knowledge and their actual performance. The construct of calibration is measured by rating an individuals' level of confidence in their ability in answering a question or recalling a piece of information and comparing to the correct answer(s). A person is said to be perfectly calibrated when his/her level of confidence corresponds to actual performance. A well-calibrated person will have a level of confidence that closely approximates actual performance and approaches perfect alignment; a poorly calibrated person shows poor alignment. The construct is considered to be a reflection of the person's learning process (Hacker & Dunlosky, 2003). Though it is well established that learners with optimal metacognitive skills accurately estimate their knowledge in a variety of domains, monitor their learning, and keep abreast of domain-specific developments (Everson & Tobias, 2000), learners exhibit confidence and generate confidence ratings from beliefs about their ability in a given domain rather than based on information presented to them during an experiment. As these studies were conducted in various domains such as physics, music, education, and history, and the conclusions appear to be consistent, it can be generalized that this will also be true for nurses who are life-long learners. The commonality among these studies is that poor calibration occurs

when learners exhibit excessive confidence based on their domain familiarity rather than knowledge gained from what is presented when making their judgements. The consequence of this poor calibration will be poor metacognitive strategies in the learning process and slow and inadequate intake of knowledge. However, the impact of poor calibration, when it occurs, will be limited to the individual only in the fields of music, history, and education. The same cannot be said for applied healthcare fields such as nursing, because nurses' knowledge and application of knowledge (cognitive tasks) influence and affect patient outcomes and safety. Kelley and Lindsay (1993) allude to a direct relationship between self-regulation and calibration. That is, well-calibrated learners are better at self-regulated learning. In addition, Stone (2000) concludes that there is a strong connection between calibration and self-regulation and states that "feedback can help students self-monitor better, which leads to more thorough self-evaluation, and hence they should become calibrated" (Stone, 2000; p439). It is evident that both calibration and self-regulated learning may tap students' motivation toward various tasks.

2.7.2 Feedback, calibration and cue-utilization

A few researchers have done a considerable amount of work in establishing the influence of feedback on individuals' calibration. Mok et al. (2015) demonstrated that feedback helps individuals' calibration in a positive way, as they become better calibrated. Benjamin and Diaz's (2008) theoretical research on the relationship between feedback and calibration, based on signal-detection theory, had the same conclusion. Several researchers found that feedback provided to a learner offers multiple benefits such as improved calibration, improved task performance, better performance

monitoring, better strategy development and deployment, and better cognitive processes (Boekaerts & Corno, 2005; Tauber & Rhodes, 2010; Wahlheim, 2011).

Stone (2000) recommends that feedback providers emphasize performance monitoring, which is a metacognitive skill, rather than actual performance. One possible reason for this recommendation is that actual performance is based on cognitive skills and knowledge and is changeable through various internal and external factors. Rhodes and Castel (2008) posited that individuals could be better calibrated on cognitive judgements if appropriate feedback were provided. This will be useful for any learner and particularly helpful to skilled learners such as nurses and other healthcare professionals, as they are often required to use cognitive judgements. In the same paper, Stone (2000, p 441) states that feedback improves calibration by bringing confidence levels into closer alignment with population norms. Although feedback on cognitive tasks may increase an individual's accuracy, a primary role of feedback in calibration is to change individuals' levels of confidence (Berger & Karabenick, 2016). That is, an individual learner may become over-confident in an increasingly complex or difficult task if the confidence level does not change or adjust, which may affect performance. Therefore, it is imperative to have the right amount of confidence, or calibration, when operating in a complex environment such as medical device alarm management.

While previous studies in education have investigated how classroom impact of other training methods, specifically one-on-one training and individualized feedback, on learners' metacognition. The literature shows that even though nurse educators have known about the concept of metacognition for several years, they have not expanded the application of it beyond the classroom in university settings (Benner et al., 2000; Billings

& Halstead, 2009). This study is the first to evaluate the following: (i) metacognition in nurses and their assistants managing medical device alarms and (ii) the influence of individualized feedback on metacognitive monitoring and control. This study will also establish the relationship between training method(s), content knowledge, and metacognition. This research will contribute to our understanding of metacognitive monitoring and control and offer instructional strategies for medical device alarm management, with an emphasis on metacognition in learning.

According to Koriat (1997), the basis of judgement of learning (JOL) and their accuracy can be explained by the cue-utilization view. This view assumes that JOL is inferential in nature: JOLs are based on the implicit application of rules or heuristics in order to achieve a reasonable assessment of the probability that the information in question will be recalled or recognized at some later time. (Koriat, 1997; p350). Koriat described three classes of information that participants may use when making JOLs: intrinsic, extrinsic, and mnemonic factors. Intrinsic factors are related to properties of the stimuli, for example, item cogency, item's reality, relatedness, and so on. Extrinsic factors are properties of the encoding conditions (e.g., study strategy, number of learning trials, etc.). Mnemonic factors refer to internal, experienced-based indicators of future recall, including memory of previous recall attempts, accessibility of target information, and cue familiarity.

2.8 Objectives

The objective of this study was to assess the effect of training methodology and feedback on two aspects of metacognition – monitoring and control. This study also hypothesized that concrete individualized feedback given during one-on-one training

sessions, based on participants' exam performance, would improve content or domain specific knowledge. Improvements in content knowledge have a significant benefit in providing critical care – reduced medical alarm sentinel events.

2.8.1 Hypotheses

Specific hypotheses investigated include:

1. Generalized feedback given in a classroom setting will not improve metacognitive prediction accuracy.
2. Interactive one-on-one training and feedback will improve nurses' metacognitive prediction accuracy.
3. Interactive one-on-one training and feedback will improve domain specific or content knowledge. That is, individual training will improve medical device alarm knowledge.
4. Generalized feedback and suggestions provided during alarm training in a class room setting will have minimal impact on study and preparation techniques thus resulting in lower exam performance.
5. Existing training methods used by hospitals do not adequately equip nurses/assistants to manage clinical alarms.

2.9 Methods

In this section, the experimental methods and participants are described to address the impact of training methods and feedback on calibration. The independent and dependent variables in this section stem from the hypotheses listed in the previous section.

2.9.1 Experimental design

An experiment with various training methods was conducted to determine if the hypotheses could be supported. This experiment used a two-way mixed model, with the exams as within-subjects factor and training methods as between-subjects factor, to analyze the nurse participants' (metacognitive) calibration scores over the three weeks alarm management course across different training methods. This helped in (i) determining any improvements in metacognitive accuracy over time and (ii) establishing the relationship between calibration scores and training over time.

2.9.2 Variables

2.9.2.1 Independent variables

The independent variables tested in this study were training type and equipment type. There were three levels for each of these variables. They are shown in Table 2.1.

Table 2.1 Independent variables and levels

Independent Variable	Levels
Training Groups	No-training group (NG) Classroom-training group (CG) Interactive one-on-one training group (OG)
Equipment	Kangaroo ® Enteral Pump Philips ® MX-40 Alaris ® 8015

Each training group was tested on all three devices listed above resulting in a 3 × 3 balanced design. These three devices were chosen based on their complexity and the author's familiarity. The Kangaroo enteral feeding pump is a low-cost and easy-to-use pump during post-operative patient care. The pump is used to deliver carbohydrates, fat,

minerals and vitamins directly into the jejunum. The Philips MX-40 is a moderately complex patient-wearable monitor and requires some training in setting up and troubleshooting. Since the device is patient-wearable, it is typically used in ambulatory care and step-down recovery in which patients are allowed to move. The Alaris pump is the most complex of all three due to its versatile nature and capability. The device can be used in any patient care setting to administer any physician-specified fluid.

It requires formal training prior to use and troubleshooting. Many hospitals develop their own handling techniques, device operation and troubleshooting protocol(s) for the Alaris pump. The schematics of these three devices are shown in Figures 2.1-2.3. Three exams were administered during the three-week alarm management-training course and performance was measured using exam scores on the three exams as mentioned earlier. Each exam consisted of 30 multiple-choice alarm and equipment operation related questions and assessed participants' understanding of content covered in training scenarios.

Figure 2.1 A schematic of Alaris® 8015 pump

Figure 2.2 A schematic of Philips ® MX-40

Figure 2.3 A schematic of Kangaroo ® enteral feeding pump

2.9.2.2 Dependent variables

The dependent variables in this study were participants' *prediction accuracy (calibration)* and *exam scores.* Calibration data were computed by asking participants prior to testing to predict scores they would get for each section and overall and by taking the difference between this predicted value (overall score is a sum of all sectional predictions) and their actual performance. Participants were directed to enter a score for each section and also an overall score at the top of the exam sheet for each exam. Overall prediction accuracy was computed using the following formula:

$$\left[1 - \frac{Prediction - Actual}{Total\ Items}\right] * 100 \qquad (3.1)$$

This formula produced a calibration score that could be expressed as a percentage and would be easier for data analysis purposes. For example, a participant who predicted a 27 on a 30-item test and earned a 29 would have a calibration score of 106.66%, miscalibrated toward under-confidence. Conversely, a participant who predicted a score of 28 but earned a 20 would have calibration score of 73.33, again miscalibrated but toward over-confidence. Nurse participants could have calibration scores ranging from 0 to 200 based on equation 2.1, where 0 and 200 indicate total miscalibration and 100 indicates perfect accuracy. A score below 100 indicates over-confidence and a score over 100 indicates under-confidence. This calibration score was computed using equation (2.1) for each participant and for all exams. Hacker et al. (2008) used the formula in 2.1 in a similar experimental study to identify over and under confidence participants. The researchers in this study used the absolute differential of predicted and obtained grades for their calibration score.

The ability of clinicians to match confidence to their judgmental abilities is a crucial link in providing patient care. Therefore, it is imperative to clearly assess and help those with both over and under-confidence. In addition to computing global prediction scores, calibration curves were plotted for global predictions against a "perfect calibration line." The exam questions used for each exam are shown in Appendix A. Each exam was a 30-question set, producing a numerical score from 0 to 30. The exam consisted of 30 multiple-choice questions on various types of alarms, working principle of the subject medical device, and practical scenarios discussed during training sessions. One point was awarded for each correct answer and no penalty was assessed for choosing a wrong answer.

2.9.3 Participants

Participants for this study included 45 Washington state licensed nurses (RNs) and CNAs who were doing clinical rotations at Seattle area hospitals. The study participants were either practicing or participating in clinical rotation programs in primary, secondary, and tertiary care settings. They were randomly assigned to one of the three training groups: no-training (NG), classroom (CG), or one-on-one (OG). The sample size was composed of 40 females and 5 males, ranging from 21 to 60 years of age with a mean of 37.06 years (SD = 10.19). Hospital research boards and word of mouth were used to recruit participants. The inclusion criterion for the study was basic proficiency in medical device alarms, assessed via an initial survey and a screening exam. There were no exclusion criteria for this study. Training sessions and administration of exams were aligned with study participants' shifts. For example, a study participant working the evening shift attended his or her assigned session and took exams in the

evening. This was to prevent collaboration and crossover communication. The control group (NG) did not go through any training. Participants from this group used the knowledge they gained through their formal education, on-the-job training, web, and self-reading to answer the exam questions. The control group was provided with user manuals for each device and were verbally told about exam score prediction and exams each week. Three types of demographic information were collected: age, clinical experience, and alarm management experience. These are summarized in the statistics section 2.11.

2.9.3.1 Estimation of sample size

Previous studies were reviewed to determine the appropriate sample size. A study by Hacker et al (2008) utilized 109 participants in assessing metacognitive monitoring and accuracy. To determine if concrete feedback helped improve metacognition, Miller and Geraci (Miller and Geraci, 2011) used 81 participants in their study. In a medical equipment alarm learnability and discriminability study, Anthony et al. (2013), used 33 psychology students. Based on a review of these studies, the sample sizes ranged from the low 30s to low 100s. Since the chosen experimental model, repeated measures ANOVA and between subjects' design, is a powerful and versatile tool, a sample size within this range could be used.

2.9.4 Experimental Protocol

The entire study took place over a course of three (3) weeks. All 45 participants, upon clearing the screening evaluation, were randomly assigned to one of the three training groups: no-training, classroom or interactive one-on-one training. All three groups were tested on all three medical devices over three weeks.

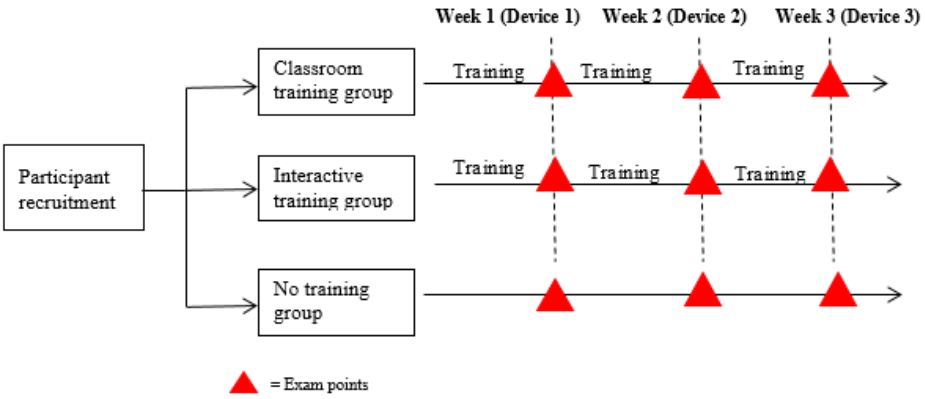

Figure 2.4 Experiment protocol and test sequence

Three medical devices were selected for this experiment based on their market penetration. Philips Monitor MX40® (Philips, Bothell, WA), Alaris Infusion Pump Model 8015® (Carefusion, San Diego, CA), and Kangaroo® Enteral Infusion Pump (Covidien, Mansfield, MA). Prior to enrolling in the class and participating in the study, all participants were be asked to complete an informed consent form approved by Mississippi State University IRB committee and a demographic questionnaire. These are shown in Appendix A.

Siebig et al. (2010) demonstrated that 92% of alarms in acute care settings are caused by infusion pumps and patient monitors; therefore, two infusion pumps and one patient monitor were chosen for this study. During the training, three areas were covered: operating techniques, various alarms, and troubleshooting for each piece of equipment. Classes took place in conference rooms for the classroom-training group and in a simulator or private meeting room for the interactive one-on-one training group.

Since the Joint Commission (a non-profit healthcare certification organization that certifies 21,000 healthcare facilities in the United States) mandated an alarm management

that program requires compliance by 2018, hospitals in the United States are in the process of formalizing procedures, and training programs and developing policies (https://www.jointcommission.org/assets/1/18/JCP0713_Announce_New_NSPG.pdf ; last accessed: September 1, 2017). Hospitals are required to document the training provided to nurses/assistants for devices containing alarms. Three hospitals in the Seattle area are currently developing a formal certification program for alarm management and planning to formally institute a program by the end of 2017. This study took place in accordance with requirements of a formal certification program and participants were a given a certification of attendance (in memo format) which they could use toward continued education credits (CEUs).

All hospital staff providing patient care using medical equipment containing alarms are required to document their training. Training provided by various private entities, for example, the device manufacturer, hospital biomedical department, or floor supervisor, is acceptable per Joint Commission guidelines. A small compensation in the amount of $25 (gift card) was also provided to participants after they completed *all* required exams. Classes were taught by the researcher with logistical help from a biomedical equipment technician (BMET)

Three exams were administered during the three-week training course, one at the end of each week. Materials covered during the week (Monday through Friday) were tested in these exams. Participants assigned to the no-training group did not receive any training but had access to manuals, brochures, and guides supplied by the device manufacturer for each device. The instruction manuals, guides, and brochures were

downloaded from the manufacturer's website for each test device and used for training purposes.

The classroom training session lasted approximately one hour each day and aligned with study participants' work shifts. The total training session time for classroom-training was five hours. The class durations were documented for each session. The one-on-one training group participants were allowed to request as many hours of training as they wanted; however, a minimum of five hours of training was required. Despite trainer's encouragement to request as many training sessions as their schedule permitted, participants (in the one-on-one training group) did not request any follow-on training sessions. Similar to the classroom group, the training duration for each session was documented. The researcher explained the score prediction (calibration) procedure to each group prior to starting the experiment. In addition, participants were also advised on ways to increase calibration over time: *participants could either adjust their prediction or they could raise their exam scores by acquiring knowledge.*

Examination questions were provided to participants on printed sheets along with a prediction sheet to each participant. Participants had five minutes to read the questions and record their score prediction onto the prediction sheet. The prediction sheets were collected prior to the start of the exam. Question papers used for all three exams are shown in Appendix A. The question paper contained three sections: basic operations (Questions 1 –10), various alarms (11 –20), and trouble shooting (21 –30). Details of class and exam(s) for the experiment are shown in Table 2.2.

Table 2.2 Class schedule

Week	**Topic**
1	Kangaroo ® Enteral Pump
	Exam 1
2	Philips ® MX-40
	Exam 2
3	Alaris ® 8015
	Exam 3

Note: (1) NG participants took the exams in the same order and on the same topics simultaneously (2) Exams correspond to the device covered during the week. i.e. Exam 1 covered Kangaroo ® Enteral Pump; Exam 2 covered Philips® MX-40 and Exam 3 covered Alaris® 8015.

Participants assigned to the classroom group received high-level generic feedback after Exam 1 and prior to starting the week 2 class. The feedback consisted of presenting the mean score and standard deviation (class) for each section, and showing the availability of study materials on various channels (websites, instructor notes, and manufacturer's printed materials). The participants were expected to use this generalized feedback to manage their time and make necessary modifications to their study habits. Both components of metacognition, metacognitive monitoring and control were expected to be affected to a certain extent for all participants, and feedback was expected to help students attain higher scores on subsequent exams. For the OG, answer sheets from Exam 1 were reviewed individually and weak areas were identified based on sectional answers. Since each section was dedicated to one area, working principle, alarms, and troubleshooting, it was easy to identify participants' weak areas. Subsequent training sessions for OG participants were based on their performance on Exam 1 and needs. For example, troubleshooting techniques was emphasized for a participant who scored low on the troubleshooting exam questions. The researcher emphasized, again after Exam 1,

ways to increase calibration overtime: *participants could either adjust their prediction or they could raise their exam scores by acquiring knowledge.* Upon completion of the training sessions and exams, all participants were asked to complete a survey questionnaire shown in Appendix A. This survey contains questions related to their formal education, training provided by the hospital and specific treatment unit, and any other co-curricular training they had received in the past. Correlation analysis was conducted between these factors and the dependent variables, prediction accuracy and exam performance.

2.10 Statistical analysis

Descriptive statistics (e.g., mean, median, and standard deviation) were computed for all dependent variables and demographics such as age, clinical experience, and alarm management experience. The demographic data summary is shown in Table 2.3. The primary analysis was a 3 (group: no-training, classroom-training, one-on-one training) × 3 (time of assessment) repeated measures ANOVA with prediction of accuracy and exam scores serving as the dependent variables. The chi-square test for independence was used to determine if there was any significant relationship between variables. Mauchly's sphericity test was used to test the assumption of compound symmetry of the common covariance matrix. As this test showed that the assumption of sphericity was violated, the lower-bound (L-B) correction method was used to determine the acceptability of hypotheses. All results were considered significant at an alpha level of 0.05. SPSS version V.25 for Windows was used for testing assumptions and Minitab R17 (for Windows) was used for other statistical analysis.

Microsoft Excel (for Windows) was used to generate graphs and assess any patterns and trends. Regression statistical models were used to analyze exam scores as a predictor of training method and device complexity (defined through exams) individually. Resulting R^2 values from the regression models were used to determine how much of the total variation in exam scores was explained by training method and level of complexity.

[Intentionally left blank]

Table 2.3 Demographics data summary

Variables	Mean (SD) or %
Age	37.1 years (10.2)
Gender	
Female	88.8%
Male	11.1%
Nursing background	
Registered nurse (RN)	37.7%
Nurse assistant (CNA)	62.2%
Years of experience in managing device alarms	
None	4.4%
Less than 1 year	4.4%
1-3 years	22.2%
3-5 years	44.4%
More than 5 years	24.4%
Training on medical device alarms?	
Yes	60.0%
No	40.0%
Adequate training provided by your institution ?*	
Yes	15.6%
No	55.6%
Did your assigned unit provide any training?	
Yes	11.1%
No	88.9%
Educational background	
CNA/other	80.0%
Associate's degree	4.4%
Bachelor's degree	15.6%
Graduate degree or higher	0.0%
Any other certifications?*	
Yes	0.0%
No	77.0%

*-Percentage does not equal 100 due to missing responses. Thirteen of the 45 study participants and 10 of the 45 study participants did not answer the questions about training adequacy and certifications, respectively.

2.11 Results

Descriptive statistics for each of the dependent variables are provided in Table 2.4. In general, exam scores increased over time for each training intervention group (CG and OG) as well as the control group (NG). As expected, one-on-one group participants' scored higher than their counterparts in the control and classroom groups. Regardless of training group, all participants' calibration scores were below 100 at Exam 1 (indicating over-confidence) and over 100 (indicating under-confidence) at exams 2 and 3. In other words, all participants exhibited over-confidence at the beginning (exam # 1) and under-

confidence during subsequent exams as they received feedback and device complexity increased.

Table 2.4 Descriptive statistics for response variables (exam score and calibration) with predictions given by participants prior to each exam

Training group	Predicted			Exam score (actual)			Calibration (calculated using formula 2.1)		
	Exam 1	Exam 2	Exam 3	Exam 1	Exam 2	Exam 3	Exam 1	Exam 2	Exam 3
No-training	23.60 (2.29)	17.07 (1.75)	19.07 (1.39)	17.13 (1.92)	18.20 (1.21)	20.07 (1.62)	78.44 (6.65)	103.78 (7.33)	103.77 (6.04)
Classroom - training	24.07 (1.81)	18.47 (2.39)	20.47 (1.81)	19.13 (1.55)	22.00 (2.17)	24.00 (1.60)	83.56 (6.84)	104.31 (4.92)	111.78 (11.88)
One-on-one training	23.53 (1.41)	19.87 (2.07)	22.13 (1.19)	20.73 (1.12)	25.93 (1.03)	26.20 (1.08)	90.67 (5.60)	120.22 (6.23)	113.56 (2.95)

Note. Average values are shown along with standard deviation within brackets.

2.11.2 Mauchly's test of sphericity assumption

As the assumption of sphericity is important in repeated measures ANOVAs, Mauchly's test of sphericity was used to test this assumption. The p-value obtained from this test is presented in Table 2.5. As the p-value > 0.05, the assumption was violated and the lower-epsilon method of correction, was used.

Table 2.5 Mauchly's sphericity test for training groups

Within subjects Effect	Mauchly's W	Approx.Chi-Square	df	Sig	Lower-bound Epsilon
Traininggroups	0.732	4.051	2.000	0.132	0.500
Exam(s)	0.865	1.887	2.000	0.389	0.500
Traininggroups* Exams	0.197	20.183	9.000	0.018	0.250

As stated, the value of parameter W was above 0.7 and 0.9 for dependent variables, which is closer to 1 and hence the assumption was met. However, Mauchly's test of sphericity indicated that the assumption of sphericity had been violated for the two-way interaction, χ^2 (9) = 20.183, p = .018, which is lower than p-value of 0.05. Therefore, the researcher interpreted the uncorrected Mauchly's W for training group and exam(s), whereas one of the corrected Mauchly's W (Epsilons) needed to be used for the interaction – training group and exam. Due to its robust nature, the Greenhouse-Geisser correction method was used when analyzing interaction.

Mauchly's test of sphericity was also performed for the second dependent variable, exam score, and presented in Table 2.6. The assumptions were met for main effects of training and exam, but were not conforming for the interaction term, χ^2 (9) = 17.494.183, p = .043. Thus, we interpreted Mauchly's W for training group and exam and one of the corrected Mauchly's W, Greenhouse-Geisser correction method for the interaction.

Table 2.6 Mauchly's sphericity test for exam

Within subjects effect	Mauchly's W	Approx. chi-square	df	Sig.	Epsilon Lower-bound
Traininggroups	0.938	.831	2.000	0.660	0.500
Exams	0.992	.103	2.000	0.950	0.500
Traininggroups * Exams	0.244	17.494	9.000	0.043	0.250

2.11.3 Calibration (prediction accuracy)

The Calibration score was found to be affected by both exam and training group. All participants exhibited over-confidence when the subject matter to be tested was easy and became under-confident as the subject matter complexity to be tested progressively

increased. Participants' calibration was below 100 at Exam 1, indicating over-confidence, and above 100, at exams 2 and 3, indicating under-confidence. Thus, exam score predictions were higher than exam scores at Exam 1, and lower exam scores at exams 2 and 3. The computed calibration at various exams is shown in Figure 2.2. As expected, the control group participants (NG) were comparatively more over confident than the classroom-training and one-on-one training group participants. In other words, one-on-one trained participants were better calibrated at Exam 1 than their counterparts. However, they were under-confident compared to their counter parts at exams 2 and 3. Factors contributing to this finding are explained in the discussion section.

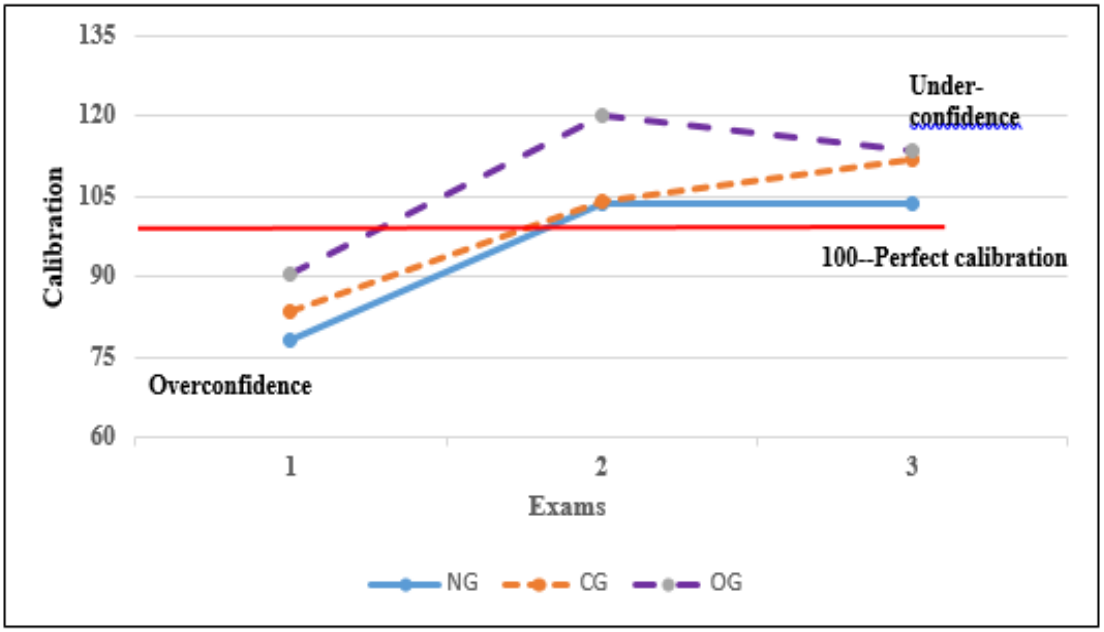

Figure 2.5 Computed calibration at various exams

NG = no-training group, CG = classroom-training group, OG = one-on-one training group.

As stated previously, participants were advised after Exam 1 and 2 on how to improve calibration. They could adjust (lower) their prediction or raise their exam scores by acquiring knowledge. In addition to the influence of device complexity, it is possible that participants adjusted their prediction, by lowering their prediction for Exam 2 and increasing their prediction for Exam 3. Regardless of the assigned training group, no significant difference was observed in the predicted scores among participants at the beginning of the study, meaning all participants gave similar predictions compared to their actual exam scores. As shown in Figure 2.3, the predictions were almost identical at the beginning (close to 24 at Exam 1) and dropped later, to a greater degree for no-training participants compared to classroom and one-on-one training group participants. In other words, prediction was highest for the one-on-one group participants and lowest for the no-training group participants at exams 2 and 3. A common result across the training groups was that all participants adjusted their predictions as time progressed: they had lowered predictions for Exam 2 and increased predictions for Exam 3.

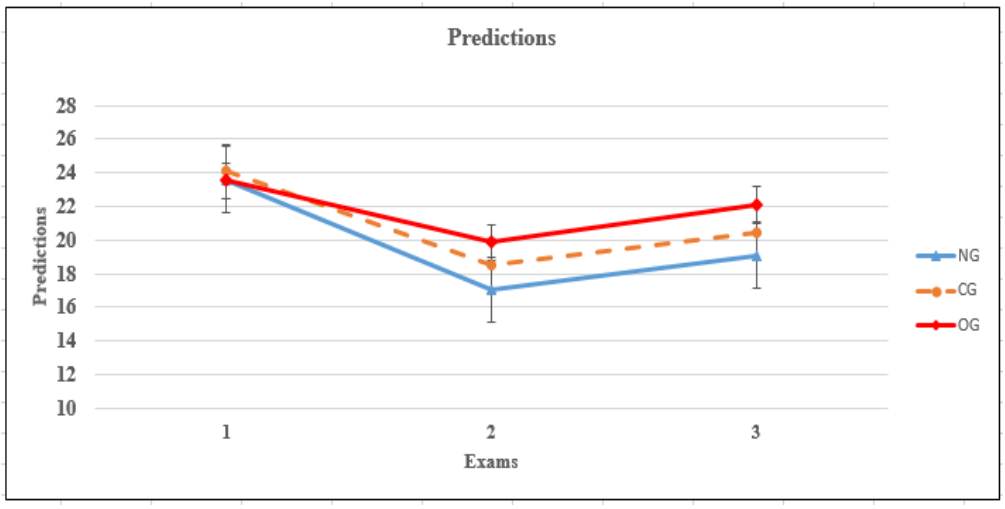

Figure 2.6 Predictions given by participants prior to exam

NG = no-training group, CG = classroom-training group, OG = one-on-one training group.

2.11.4 Exam performance

Figure 2.4 shows the trend of actual exam scores obtained by participants. The effect of training was found to be significant from the beginning to end. Though the difference in exam scores between training groups was smaller at the start, it was remarkably higher by Exam 2 for the classroom and one-on-one training groups compared to the no-training group, and remained higher until the end of the study for the one-on-one trained participants and continued its increasing trend for the classroom participants. The exam score plateau observed for the one-on-one group participants may have been due to a "ceiling effect," in that the maximum score had been attained for their potential and hence very little room was left for further improvement. Irrespective of the training method, all participants showed improvement in scores over time. Compared to Exam 1, no-training group (NG) participants showed a small incremental improvement in exams 2 and 3, whereas classroom (CG) and one-on-one (OG) training group participants scored significantly higher on exams 2 and 3 than Exam 1. It is important to note that participants improved their scores over time despite the progressive device complexity increase in exam(s). Based on the graph, one can conclude that training may not be a significant factor for lower complexity devices, whereas it could play a key role in educating nurses and caregivers when the medical device is complex.

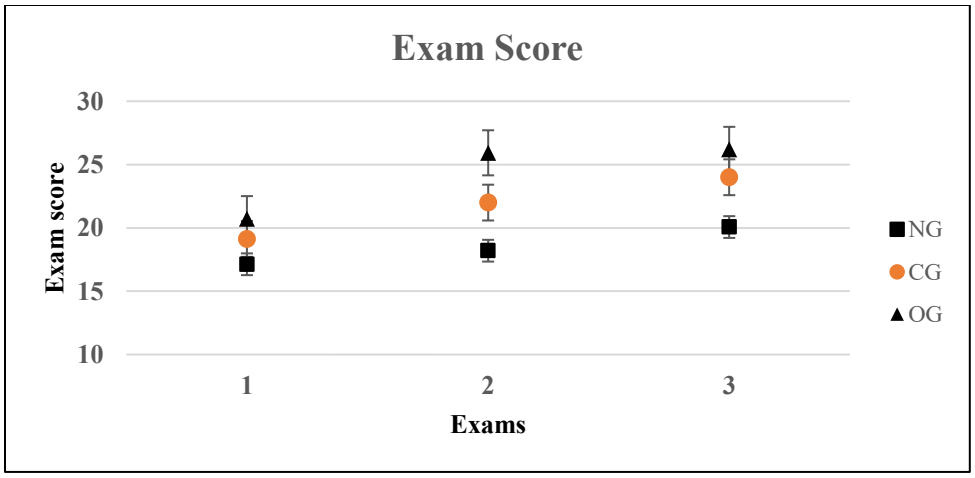

Figure 2.7 Exam scores across three exams for training groups

NG = no-training group, CG = classroom-training group, OG = one-on-one training group.

2.11.5 Analysis of variance (ANOVA) for calibration

To determine whether training methodology and exam(s) affected calibration (metacognitive prediction accuracy), students' calibration scores on the three exams were analyzed using repeated measures ANOVA and the main effects of training method and exams were reviewed. The reliability analyses indicated an acceptable level of reliability between exams ($\alpha = .70$). Results revealed that overall, both training method and device complexity, which was assessed through exams, were significant in determining calibration. The main effect of training group showed a statistically significant difference, $F(2, 28) = 25.876$, $p < 0.05$, the main effect of exam showed a statistically significant difference, $F(2, 28) = 231.495$, $p < 0.05$, and the R^2 (adj) value for the model of 76.63% indicated adequate fitness. ANOVA results are shown in Table 2.7.

Table 2.7 ANOVA output for calibration

Source	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Traininggroups	3876.738	1.000	3876.738	25.876	.000	.649
Error(Traininggroups)	2097.447	14.000	149.818			
Exam(s)	19164.071	1.000	19164.071	231.495	.000	.943
Error(Exams)	1158.975	14.000	82.784			
Traininggroups * Exam	767.984	1.000	767.984	4.871	.045	.258
Error(Traininggroups* Exams)	2207.255	14.000	157.661			

As shown in Table 2.7, the interaction was statistically significant, F (1, 14) = 4.871 at a p-value of 0.045; thus, it is important to examine the post-hoc pairwise comparisons to assess the influence of training method along with device (s). Pairwise estimates between training groups across exam(s) are shown in Table 2.8. At Exam 1 which covered Kangaroo® Enteral Pump (beginning of the experiment), the calibration score for the no-training group was 78.445 (95% CI 74.761 and 82.128) and for the classroom-training group was 83.555 (95% CI 79.767 and 87.342), which was not a statistically significant difference. However, it was 90.667 (95% CI 87.455 and 93.879), for the one-on-one trained participants indicating that they were better calibrated than their counterparts. That is, the closer the calibration score is to a perfect calibration line of 100, the better it is from a prediction accuracy standpoint. With an increase in device complexity, one-on-one trained participants erred toward under-confidence: their calibration scores were 120.223 (95% CI 116.771 and 123.674) and 113.556 (95% CI 111.924 and 115.188) at exams 2 and 3, respectively. It is possible that the feedback provided was salient and participants were exposed to more technical knowledge than they needed to know, which led them to become under-confident. It is plausible that substantive (device-specific) individualized feedback provided as part of one-on-one

training is related to the formation of metacognitive judgments, while calibration feedback provided during classroom training involves the translation of those judgments into overt knowledge. Classroom training allowed participants to gain knowledge more than that of the no-training group in that all pairs were significant at a value of $p < 0.005$. Therefore, it is reasonable to conclude that current training methods, such as peer-to-peer, online, manufacturer-offered, and on-the-job are not effective in gaining alarm related knowledge. At the very least, training should be classroom-based with a smaller group to allow for more engagement and feedback tailored to the needs of the group.

[Intentionally left blank]

Table 2.8 Pairwise estimates of calibration for training groups

Traininggroups	Exam	Mean	Std. error	95% Confidence Interval Lower bound	Upper bound	Partial eta^2	t
1	1	78.445	1.717	74.761	82.128	.993	45.680
	2	103.778	1.893	99.718	107.838	.995	54.823
	3	103.333	1.560	99.987	106.678	.997	66.241
2	1	83.555	1.766	79.767	87.342	.994	47.312
	2	104.311	1.271	101.586	107.037	.998	82.091
	3	111.777	3.067	105.200	118.355	.990	36.449
3	1	90.667	1.498	87.455	93.879	.996	60.539
	2	120.223	1.609	116.771	123.674	.997	74.713
	3	113.556	.761	111.924	115.188	.999	149.222

Table 2.9 Pairwise comparisons for calibration

(I) Traininggroups	(J) Traininggroups	Mean difference (I-J)	Std. error	Sig.	95% Confidence Interval for Difference[a] Lower bound	Upper bound
1	2	-4.696	2.224	.160	-10.741	1.349
	3	-12.963*	1.417	.000	-16.813	-9.113
2	1	4.696	2.224	.160	-1.349	10.741
	3	-8.267*	1.742	.001	-13.002	-3.533
3	1	12.963*	1.417	.000	9.113	16.813
	2	8.267*	1.742	.001	3.533	13.002

Based on estimated marginal means
* - The mean difference is significant at the 0.05 level. [a] – Adjustment for multiple comparisons: Bonferroni

2.11.6 ANOVA for exam performance

The repeated measures ANOVA was conducted using exam score as the dependent variable to assess the impact of training methodology and device complexity. The reliability analyses indicated a good level of reliability between exams (α = .70). Results revealed that overall, both training method and exam(s), were significant in determining exam scores. The main effect of training showed a statistically significant

difference, F (2, 28) = 126.64, $p < 0.05$, the main effect of exams showed a statistically significant difference, F (2, 28) = 168.624, $p < 0.05$, and R^2 (adj) value for the model of 76.44% indicated adequate fitness. The ANOVA results are shown in Table 2.10.

Table 2.10 ANOVA for exam score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Traininggroups	766.993	1.000	766.993	126.648	.000
Error(Traininggroups)	84.785	14.000	6.056		
Exam	468.637	1.000	468.637	167.624	.000
Error(Exam)	39.141	14.000	2.796		
Traininggroups * Exam	71.319	1.000	71.319	7.020	.019
Error(Traininggroups* Exam)	142.237	14.000	10.160		

The parameter estimates of exam scores for each training group across different exams are shown in Table 2.11. As expected, the no-training group scored the lowest and the one-on-one training group scored the highest; classroom-trained participants scored in between. Post-hoc analysis with a Bonferroni adjustment revealed a difference in exam score from 18.567±0.27 on Exam 1 to 21.75±0.27 on Exam 2, a statistically significant increase of 3.28 (95% CI, 2.21 to 4.36), $p < .0005$, and an increase in exam score from 21.75 on Exam 1 to 24.28 on Exam 2, a statistically significant increase of 2.53 (95% CI, 1.49 to 3.57), $p < .0005$, clearly demonstrating the effect of training.

Table 2.11 Parameter estimates for exam scores for each training group

Trainninggroups	Mean	Std. Error	95% Confidence Interval Lower Bound	Upper Bound
1	18.467	.278	17.871	19.062
2	21.756	.273	21.170	22.341
3	24.289	.169	23.927	24.651

Table 2.12 Pairwise comparison for each exam scores for each training group

(I) Trainninggroups	(J) Trainninggroups	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[a] Lower Bound	Upper Bound
1	2	-3.289*	.394	.000	-4.360	-2.217
	3	-5.822*	.319	.000	-6.688	-4.957
2	1	3.289*	.394	.000	2.217	4.360
	3	-2.533*	.383	.000	-3.575	-1.492
3	1	5.822*	.319	.000	4.957	6.688
	2	2.533*	.383	.000	1.492	3.575

Based on estimated marginal means
* –The mean difference is significant at the .05 level. [a]–Adjustment for multiple comparisons: Bonferroni..[b] – substantially less than significance level.

2.12 General discussion

The central question of this experiment was whether training and feedback methods have an effect on individuals' judgment and performance. Based on the data collected on two dependent variables, calibration and exam performance, the hypothesis was supported. The degree to which training and feedback have an effect is explained as follows.

Nurses who received one-on-one training and personalized feedback performed better than their untrained and classroom-trained counterparts. Although training and feedback improved one-on-one trained participants' performance (better exam scores),

their calibration bias did not improve. In the current study, students who engaged in confidence estimation started the course over-confident and ended the course under-confident, regardless of training group. That is, all participants started Exam 1 over-confident and completed Exam 3 under-confident. At Exam 1, no significant difference was observed in calibration among the training groups; however, by Exam 2 and at Exam 3, one-on-one trained participants were poorer in calibration than their counterparts in the classroom and control groups: they were highly under-confident compared to their counterparts. These findings converge with those of Callender et al. (2016). Researchers in that study observed under-confidence behavior among higher performers and feedback influenced calibration to a greater extent than they anticipated.

The results in the current study suggest that training on concepts of metacognition and calibration directly affects performance. Although the feedback provided to classroom participants was generic in nature, it mainly targeted accuracy of participants' judgements and included ways to improve calibration. Thus, it essentially served as "performance" feedback. In addition to this performance feedback, one-on-one participants also received feedback on areas in which they underperformed. This type of feedback provided to one-on-one participants on device-specific matters was "environmental" in nature. Environmental feedback refers to subject-specific information given about the task under consideration. This type of feedback increases one's substantive (domain-specific) expertise, which was the study's intent. As such, receiving repeated and targeted environmental feedback makes one a subject matter expert. It was anticipated that environmental feedback would not only improve substantive knowledge, but also would help improve metacognitive accuracy. The results obtained do not support

this premise. It is possible that one-on-one participants absorbed the environmental feedback to grasp the fundamentals and relied solely on it when making predictions. Their predictions were very similar to other participants in the beginning (over–confident); however, at exams 2 and 3, they scored much higher than what they predicted, thus showing under-confidence. The findings from this experiment converge with those of Stankov and Crawford (1997), who found that higher performers are generally under-confident compared to lower performers, who tend to be over-confident irrespective of the difficulty of subjects and tasks. Further, they are also in alignment with Smith and Dumont's (1997) study outcomes, which were directly applicable to clinical psychology and medical fields.

Results in the current study suggest that one-on-one training and feedback increased only substantive expertise, and not metacognitive prediction accuracy. This could be further explained using cue-utilization view, a well-established concept in metacognition, which assumes that judgments of learning (JOLs) are based on inferences from mnemonic cues inherent to learning process. It is well known that metacognitive monitoring capitalizes on correlations in the "internal ecology" of cognitive processes between mnemonic cues and actual memory. In a cue-utilization study by Koriat (1997), the calibration of study participants was impaired by overlearning and repeated exposures of study material whereas the performance, assessed through testing, had improved. The study participants in our experiment, discounting extrinsic cues presented to them through one-on-one training and feedback, exhibited under-confidence. They were comparatively more under-confident than their control group peers. On the other hand, control group and to a certain extent classroom-trained participants, used their internal

mnemonic cues in predicting their scores, which found to be closer to the perfect calibration than one-to-one trained participants. Mnemonic cues are found to be based on many factors such as task experience, knowledge acquired over time and task exposure among others. These knowledge based cues likely helped control group participants in making better predictions whereas the prediction-based cues, reinforced through feedback and one-to-one training did not help in prediction to the extent we desired. Results from the current experiment concurs with results from Koriat's cue-utilization study. In general, predictions (judgements of learning) are accurate as long as the cues used at the time of making such predictions are consistent with the factors that affect performance later. According to Koriat, "the increased reliance on mnemonic cues with practice may be expected to improve judgement of learning accuracy because such cues reflect the effects of past experience and can serve as a good basis for memory predictions (Koriat, 1997; p349). To summarize, in making judgments of learning, participants do not monitor directly the strength of the memory trace of the item in question, but use a variety of cues that are generally predictive of subsequent memory performance.

Based on the results of this study, neither excessive substantive training beyond what is needed to address clinical conditions nor enriched external cues may not improve metacognitive accuracy. One-on-one trained participants were not better calibrated, and erred toward under-confidence. Under-confidence in clinicians is not necessarily harmful. Under-confidence may, however, be the preferable error in a clinical setting: Nurses who are over-confident when an alarm goes off may prepare inadequately for such situations and make mistakes, while nurses who act based on under-confidence may be less likely to do so and exercise caution before taking action. There are many benefits for being

under-confident. Clinicians who believe they do not know what they need to know to reach a decision, will be highly motivated to seek additional knowledge and receptive to any knowledge or suggestions from external sources. They will not resort to trial-and-error type learning method. As under-confident clinicians will attempt to obtain additional resources and allocate appropriate time to diagnose and resolve medical alarms, the potential for adverse events will be reduced resulting in enhanced patient safety and better healthcare environment.

As explained previously, feedback provided to classroom participants, though generic in nature, helped them secure better overall calibration scores compared to one-on-one participants; however, their exam scores were lower across all exams. In other words, classroom participants' calibration accuracy was superior and exam performance was inferior comparing to one-on-one trained participants. Surprisingly, their calibration scores were not significantly different from control group participants for exams 1 and 2 and seemed to be more under-confident at Exam 3. This is likely because the syllabus covered for Exam 3 was difficult, and they could have been attentive during classroom sessions and used appropriate control strategies in preparing for the exam. Hence, they knew more than they thought they did, thus resulting in under-confidence and better exam performance. It is important to note that despite of increase in device complexity (simpler device in week 1 and a complex pump in week 3) and different set of questions in each exam, participants' calibration was getting closer to the 'perfect' calibration line in exam 3, across all groups. Though there is no learning effect due to different questions and devices tested in each exam, there may be small amount of practice effect. That is, participants, after going through two exams, grasped the central theme of questions by

exam 3 and better aligned their judgement/expectations resulting in better calibration score. Repeated measurement of calibration likely contributed to better prediction by exam 3. It is well known that participants perform better on items to which they were exposed to them few times than items that are presented a single time (McDaniel et al, 2007).

2.12.1 Test hypotheses

Hypothesis 1: Generalized feedback given in a classroom setting will not improve metacognitive prediction accuracy.

This hypothesis was well supported. Post hoc analysis with a Bonferroni adjustment revealed that calibration score was not statistically significantly increased from a non-trained control group to classroom trained group [M = 4.696 (95% CI -1.349 to 10.741); p = .160]. Generalized feedback given during classroom training sessions and after exams did not improve metacognitive prediction accuracy. Consistent with previous findings in fields such as psychology and education, generalized feedback to improve performance related to metacognitive prediction.

Hypothesis 2: Interactive one-on-one training and feedback provided will improve nurses' metacognitive prediction accuracy.

This hypothesis was not supported. There was a statistically significant difference in mean calibration from 95.18±1.43 for the no-training group to 108.14±0.62 for the one-on-one trained group, a difference of 12.96 (95% CI, 9.11 to 16.81), p < .0005. Surprisingly, the one-on-one trained participants' calibration did not improve during the experiment. It is possible that over-training exposed them to more alarm and device related knowledge than they needed to know for everyday operation. This exposure,

though it improved their domain-specific knowledge, did not help improve metacognitive prediction accuracy. Similar findings have been reported by other researchers over a semester long psychology class (Miller & Geraci, 2011; Miller & Geraci, 2016). It is important to note that the metacognitive prediction accuracy dropped from Exam 2 to Exam 3, showing a small incremental improvement, yet in the right direction, towards perfect calibration. At the beginning (Exam 1), the OG exhibited over-confidence (M = 90.66, SD = 5.60) and upon receiving tailored training and feedback on their performance in Exam 1, they swung to under-confidence at Exam 2 (M = 120.22, SD = 6.23) and stayed under-confident until the end of Exam 3 (M = 113.56, SD = 2.95). It is worth emphasizing that although this research demonstrated that substantive knowledge and calibration were altered through different routes, it did not demonstrate precisely what those routes are and why learners trained through one-on-one training were able to improve their alarm knowledge but not metacognitive calibration. It is possible that learners, as explained before, predicted based on what they did not know or were not aware of, rather than what was known already.

Hypothesis 3: Interactive one-on-one training and feedback will improve domain specific or content knowledge. That is, individual training will improve medical device alarm knowledge.

This hypothesis was well supported. Post-hoc analysis with a Bonferroni adjustment revealed that exam score differed significantly between a non-trained control group and a one-on-one trained group [M = 5.822 (95% CI -4.957 to 6.688); $p < 0.0005$]. Individualized training and tailored feedback given during one-on-one sessions and after exams significantly improved alarm-related knowledge. The results demonstrate that

individualized environmental feedback, provided based on learners' strengths and weaknesses, can change how they perform by providing an indication that additional study time is warranted. Hence, a learner appropriately plans and uses self-regulated strategies to acquire knowledge.

Hypothesis 4: Generalized feedback and suggestions provided during alarm training in a classroom setting will have minimal impact on study and preparation techniques thus resulting in lower exam performance.

Results for this study supported this hypothesis. As shown in Figure 2.6, the generalized feedback and suggestions provided to classroom participants impacted metacognitive monitoring and control strategies minimally. Further, a two-sample t-test was performed, which showed that the classroom-trained participants scored significantly lower after receiving only generalized feedback compared to one-on-one trained participants, $t(87) = -4.25$, $p < .0005$. These findings coincide with other research reporting that classroom-trained participants, who invest less effort in tasks, show less active engagement in the process of learning and self-regulation than students who invest more effort (one-on-one training). It is possible that classroom-trained participants' awareness about available resources was lower and did not know how to regulate their engagement with the subject (exam) to get better scores. That is, they relied only on classes and generalized feedback provided to them, which were not adequate to gain knowledge. In other words, classroom training and feedback did not help in improving self-regulation (preparation techniques), which were already low. Though improvements were noted between exams (mean score at Exam 1 = 19.13; at Exam 2 = 22.00; and at

Exam 3 = 24.13), scores were still lower than those of one-on-one participants at every exam (mean score at Exam 1 = 20.73; Exam 2 = 25.93; and at Exam 3 = 26.20).

Hypothesis 5: Existing training methods used by hospitals do not adequately equip nurses/assistants to manage clinical alarms.

This hypothesis was well supported. The main effect of training in the ANOVA showed a statistically significant difference for exam scores, $F(2, 28) = 25.876$, $p < 0.0005$. The entire ANOVA table is shown in Table 2.8. As shown in Figure 2.8, the control group participants had the lowest score across all exams. Results from this study show that existing training methods used by nurses, such as online material posted by manufacturers, printed training materials supplied manufacturers and hospital biomedical departments, on-the-job training by supervisors or mentors, training by peers, and sporadic classroom training are not helpful in addressing alarms that occur every day for patients. Two participants shared the following: “I don’t receive any structured training. If we received structured training, we would do better.” Hospital biomedical and training departments should take note of this and develop appropriate training programs to reduce sentinel and adverse events and improve patient safety.

2.13 Conclusions

The purpose of this study was to determine how training should be conducted to improve the accuracy of judgments and alarm knowledge of professional learners, such as nurses. This study showed that different types of training are required to improve different aspects of metacognitive prediction accuracy and alarm-related knowledge. Based on this research, individualized one-on-one training and device specific feedback

appear to be necessary for improvements in alarm knowledge, while calibration and generalized feedback is needed for better metacognition.

Although both types of miscalibration—over-confidence and under-confidence—are not desirable in any field, the most often characterized metacognition characteristic is over-confidence, as an over-confident person performs poorly and often fails to recognize their own performance level. Kruger and Dunning (1999) called these individuals "unskilled and unaware" and labeled them as "doubly-cursed." This is unsettling and unacceptable, especially in a field such as nursing because of its impact on patient outcomes. Therefore, it is important to clearly delineate under what circumstances nurses exhibit over- and under-confidence, and what roles training and feedback play. One measure to assess an individual's confidence level is calibration score. This study demonstrated that (i) by individualizing training, we can improve a learner's performance but not their calibration (judgement/prediction accuracy); (ii) one-on-one training, though expensive to implement, is a good teaching method to impart knowledge on complex subjects such as medical device alarms and alarm management; and (iii) nurses and their assistants did not or could not improve calibration through classroom teaching and one-on-one teaching methods.

Future research in improving metacognitive monitoring and control during one-on-one training and classroom training needs to be conducted. There are multiple ways to achieve this. For example, quizzes and spot-tests (between exams) can be part of the training plan, which help adjust calibration continuously. This study showed that feedback on performance and calibration in one-on-one training allowed those who

initially failed to identify why they failed and to gain a deeper understanding of various alarms, thus resulting in improvements on later exams.

The findings from this study are consistent with existing research that has shown the ineffectiveness of generalized feedback in improving metacognitive prediction accuracy. Interestingly, this study shows that the classroom teaching method improves substantive (domain- specific) knowledge more than previously thought. This may be because students are also full-time working professionals and subjects taught were directly applicable to their everyday work. Therefore, based on the findings from this study, appropriate methods should be used when designing training programs. For example, classrooms should be used to teach about moderately complex devices and alarms with multiple sessions that repeat topics, and one-on-one training methods should be used to teach about overly complex devices, as this may accelerate the substantive (device-specific) knowledge acquisition process. This study clearly showed that existing methods used by hospitals such as online training over the intranet, on-the-job training, and the "buddy system" (peer-to-peer) training do not adequately equip nurses to manage alarms as the control group participants scored the lowest. This study also helps us to conclude that regardless of how much training we provide, metacognition is somewhat resistant to significant improvements in professionals like nurses. This is because learners will exercise excessive caution when they give predictions or may be reluctant to give higher predictions, as they do not want to come across as presumptuous to their peers; thus, they will tend to predict moderately high.

2.14 Limitations and future research

The population sampled in this study needs to be expanded to include additional work groups, such as biomedical technicians, doctors, unit technicians, and laboratory personnel who help address alarms within the hospital. The experiment was conducted using three therapeutic devices (that administer physician-specified fluids or monitor patient conditions); however, alarms are ubiquitous in a typical hospital. Therefore, future research should include diagnostic equipment, laboratory appliances, analyzers, and transportation equipment. One other important limitation in this study is the number of times the students were assessed. The current study stopped at three exams, when the one-on-one training group started showing improvement in metacognitive monitoring and control. Recall that one-on-one trained participants' calibration scores started dropping from Exam 2 to Exam 3 towards perfect calibration of 100. Had the study included a few more assessment time points, whether they would have reached perfect calibration could have been verified. Systematically manipulating feedback in the one-on-one training method would provide valuable information about how students respond to environmental feedback in terms of study habits (both strategies and implementation) and the resulting performance and prediction accuracy.

During this experiment, only two dependent variables, calibration and exam performance, were considered. These two do not give a complete picture of metacognition in nurses. Other variables such as scatter, discrimination, and a relative calibration index should be studied to characterize metacognition completely. The conditions of the experiment also need to be controlled tightly. For example, during the one-on-one training session, some participants attended training sessions immediately

after caring for patients or addressing alarms. They tried to connect the class material to a situation and made the training substantive (domain–specific) than it already was. In terms of other limitations, it was also difficult to control when participants took the test. Some participants attended the training session in the morning and took the test in the evening thus allowing only few hours in between whereas other participants attended their last training session in the evening (after their shift) and took the test next evening prior to starting their shift thus allowing close to 24 hours. As participants generally tend to have good retention of material immediately after training session, taking test on the same day will likely result in better exam performance. Some participants from both experimental groups took the test on the same day as their last training session. Participants who expressed interest in this study were working professionals and their schedules had to be accommodated. Future studies could include 'cooling off' period between training and exams.

Future researchers are encouraged to have a strict protocol or script for bringing these participants back to the topics to be discussed and metacognition. As it is not economical to train every clinician in the hospital individually on every device, nurse educators and the biomedical training department can choose one-on-one training methods for certain devices based on complexity and user population, etc. They can also use specific teaching strategies and impart skills in training sessions that can be adapted to classroom contexts. This will help in reducing the cost burden associated with one-on-one training. Given that the study took place in the state of Washington, participants were employed in local area hospitals. The protocols, practice, and methods for medical device alarms are dependent on the hospital and its workforce culture. Therefore, before the

outcomes of this study are generalized, the experiment should be broadened with a diverse and large sample population.

[Intentionally left blank]

2.15 References

Alkan, F., & Erdem, E. (2012). The relationship between teacher self-efficacy and competency perceptions of chemistry teacher candidates. *Procedia - Social and Behavioral Sciences*, 47, 1927–1932.

Allen, B. A., & Armour-Thomas, E. (1991). Construct validation of metacognition. *Journal of Psychology*, 127(1), 203–211.

Anderson, N. J. (2003). Metacognitive reading strategies increase L2 performance. *The Language Teacher*, 27, 20-22.

Anthony, J. S., Clayton, K. E., & Zusho, A. (2013). An investigation of students' self-regulated learning strategies: Students' qualitative and quantitative accounts of their learning strategies. *Journal of Cognitive Education and Psychology*, 12, 359–373.

Batang, B. L. (2015). Metacognitive strategy awareness and reading comprehension of prospective pre-service secondary teachers. *Asia Pacific Journal of Multidisciplinary Research*, 3(4), 62–67.

Begley, C., & White, P. (2003). Irish nursing students' changing self-esteem and fear of negative evaluation during their pre-registration program. *Journal of Advanced Nursing*, 42(4), 390–401.

Benjamin, A. S., & Diaz, M. (2008). Measure of relative metamnemonic accuracy. In J. Dunlosky & R. A. Bjork (Eds.), Handbook of memory and metamemory (pp. 7394-7395). New York, NY: Psychology Press.

Benner, P., Sutphen, M., Leonard, V., & Day, L. (2010). *Educating nurses*. San Francisco, CA: Jossey-Bass.

Berger, J.-L., & Karabenick, S. A. (2016). Construct validity of self-reported metacognitive learning strategies. *Educational Assessment*, 21, 19–33.

Bielaczyc, K., Pirolli, L. P., & Brown, L. A. (2009). Training in self-Explanation and self-Regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction*, 13(2), 195-206.

Billings, D. M., & Halstead, J. A. (2009). Teaching in nursing: A guide for faculty. (pp 65-78). St. Louis, Missouri: Saunders.

Boekaerts, M. & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology: An International Review*, 54, 199–231.

Bol, L., & Hacker, D. J. (2001). A comparison of the effects of practice tests and traditional review on performance and calibration. *The Journal of Experimental Education*, 69, 133–151.

Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *The Journal of Experimental Education*, 73, 269–290.

Brookhart, S. M. (2008). How to give feedback to your students (pp. 47–58). Alexandria, VA: Association for Supervision and Curriculum Development.

Callender, A., Franco-Watkins, M. A., & Roberts, S. A. (2016). Improving metacognition in the classroom through instruction, training, and feedback. *Metacognition Learning,* 11, 215 –235.

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education*, 4, 404-418.

Chua, R. Y. J., Morris, M. W., & Mor, S. (2012). Collaborating across cultures: Cultural metacognition and affect-based trust in creative collaboration. *Organizational Behavior and Human Decision Processes*, 118, 179 –188.

Clarke, N. (2012). Evaluating leadership training and development: A levels-of-analysis perspective. *Human Resource Development Quarterly*, 23, 441–460.

Conway, M. A. (2005). Memory and the self. *Journal of Memory and Language*, 53, 594–628.

Coutinho, S. A. (2007). The relationship between goals, metacognition, and academic success. *Educate*, 7(1), 39-47

de Bruin, A. B. H., & van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction*, 22, 245–252.

Desoete, A. (2007). Evaluating and improving the mathematics teaching-learning process through metacognition. *Electronic Journal of Research in Educational Psychology*, 5, 705–730.

Desoete, A., & Roeyers, H. (2006). Metacognitive macroevaluations in mathematical problem solving. *Learning and Instruction*, 16, 12–25.

Duffy, K., & Hardicre, J. (2007). Supporting failing students in practice: Assessment. *Nursing Times*, 103(47), 28-29.

Duncan, T. G., & McKeachie, W. J. (2005). The making of the motivated strategies for learning questionnaire. *Educational Psychologist*, 40(2), 117-128.

Dunlosky, J., Rawson, K. A., Marsh, E. J., & Nathan, M. J., et al. (2013). Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14, 4–58.

Ellis, D. M. (2016). The role of nurse educators' self-perception and beliefs in the use of learner- centered teaching in the classroom. *Nurse Education in Practice*, 16, 66–70.

Ewing, B. (2005). A model for analysing one-to-one teaching in the maths recovery program (pp 35–59). Flaxton, AUS: PostPressed.

Farrell, K., Payne, C., & Heye, M. (2015). Integrating interprofessional collaboration skills into the advanced practice registered nurse socialization process. *Journal of Professional Nursing*, 31, 5–10.

Fawcett, J. (2014). Thoughts about interprofessional education [essays on nursing science]. *Nursing Science Quarterly*, 27, 178–179.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist*, 34, 906–911.

Gao, M., & Quitadamo, I. J. (2015). Using inquiry teaching to promote student critical thinking and content knowledge in a nonmajors biology course. *Education Journal*, 4(4), 182–188.

Good Practice. (2013). Giving feedback using the BOOST Model. tinyurl.com/a8wfeq. Last accessed: July 2, 2017.

Govaerts, M. J. B. (2008). Educational competencies or education for professional competence? *Medical Education*, 42, 234–236.

Graesser (Eds.), Metacognition in educational theory and practice (pp. 1–23). Hillsdale, NJ: Erlbaum.

Gray, M. A., & Smith, L. N. (2000). The qualities of an effective mentor from the student nurse's perspective: findings from a longitudinal qualitative study. *Journal of Advanced Nursing*, 32(6), 1542–1549.

Greene, J. A., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologists*, 45, 203–209.

Gutierrez, P. A., & Schraw, G. (2015). Effects of strategy training and incentives on students' performance, confidence, and calibration. *The Journal of Experimental Education*, 83(3), 386– 404.

Hacker, D. J. (1998). Definitions and empirical foundations. In D. J. Hacker, J. Dunlosky, & A. C. Greasser. The educational psychology series. Handbook of metacognition in education (pp. 1-23). New York: Routledge/Taylor & Francis Group.

Hacker, D. J., & Dunlosky, J. (2003). Not all metacognition is created equal. *New Directions for Teaching and Learning*, 95, 73–79.

Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: the effects of incentives, reflection, and explanatory style. *Metacognition and Learning*, 3, 101–121.

Hacker, D. J., Bol, L., Horgan, D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, 92, 160–170.

Hamid Y., & Mahmood, S. (2010). Understanding constructive feedback: a commitment between teachers and students for academic and professional development. *Journal of Pakistan Medical Association*, 60(3), 224–227.

Huang, J. S., Yang, J. H. S., Chiang, Y. F. P., & Tzeng, S. Y. L. (2012). Building an e-portfolio learning model: Goal orientation and metacognitive strategies. *Knowledge Management & E-Learning: An International Journal*, 4(1), 2073-2081.

Ibabe, I., & Juaregizar, J. (2010). Online self-assessment with feedback and metacognitive knowledge. *Higher Education*, 59, 243 – 258.

Jeffries, P. R. (2012). Simulation in nursing education: From conceptualization to evaluation. New York: National League for Nursing.

Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the test or testing to teach? Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology Review*, 26, 307–329.

Kelemen, W. L., Frost, P. J. & Weaver, C. A. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition*, 28 (1), 92-104.

Kelley, C. M., & Lindsay, S. D. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, 32, 1–24.

Koh, L. C. (2007). Refocusing formative feedback to enhance learning in pre-registration nurse education. *Nurse Education in Practice*, 8(4), 223–230.

Koriat, A., (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgements of learning. *Journal of Experimental Psychology: General*, 126(4), 349-370.

Kruger, J., & Dunning, D. (1999). Unskilled and Unaware of It: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134.

Kuiper, R., & Pesut, D. (2004). Promoting cognitive and metacognitive reflective reasoning skills in nursing practice: Self-regulated learning theory. *Journal of Advanced Nursing*, 45(4), 381−391.

Leybina, V. A., & Skvortsova, V. Y. (2009). Metacognitive component of work performance. Proceedings of the 2009 International Association of Computer Science and Information Technology - Spring Conference. 277-280.

Lunenberg, M. (2010). Characteristics, scholarship and research of teacher educators. In B. McGraw, P. L. Peterson, & E. Baker (Eds.), International encyclopedia of education (pp. 676- 680). Oxford, UK: Elsevier.

Lunney, M. (2008). Current knowledge related to intelligence and thinking with implications for the development and use of case studies. *International Journal of Nursing Terminologies and Classifications,* 19(4), 158−162.

Maki, R. H., Foley, J. M., Kajer, W. K., & Thompson, R. C. et al. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 609–616.

Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology*, 97, 723– 731.

McAllister, M., Reid-Searl, K., & Davis, S. (2013). 'Who is that masked educator?: Deconstructing the teaching and learning processes'. *Nurse Education Today*, 33 (12), 1453–1458.

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., & McDermott, K. B., et al. (2013). Quizzing in middle-school science: successful transfer performance on classroom exams. *Applied Cognitive Psychology*, 27, 360–372.

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, 19, 494-513.

Melby-Lervag, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, 49, 270–291.

Memnun, D. S., & Hart, L. C. (2012). Elementary school mathematics teacher trainees' metacognitive awareness levels: Turkey case. *Journal of International Education Research*, 8(2), 173–182.

Miller, T. M., & Geraci, L. (2011). Training metacognition in the classroom: the influence of incentives and feedback on exam predictions. *Metacognition Learning*, 6, 303–314.

Miller, T. M., & Geraci, L. (2011). Unskilled but aware: Reinterpreting overconfidence in low- performing students. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 37, 502–506.

Miller, T. M., & Geraci, L. (2014). Improving metacognitive accuracy: How failing to retrieve practice items reduces overconfidence. *Consciousness and Cognition*, 29, 131–140.

Miller, T. M., & Geraci, L. (2016). The influence of retrieval practice on metacognition: The contribution of analytic and non-analytic processes. *Consciousness and Cognition*, 42, 41–50.

Mok, M. M. C., McInerney, D. M., Zhu, J., & Or, A. (2015). Growth trajectories of mathematics achievement: Longitudinal tracking of student academic progress. *British Journal of Educational Psychology*, 85, 154–171.

Mokhtari, K., & Reichard, C. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology*, 94, 249–259.

Nicol, J. D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 2, 199-218.

O'Malley, J. M. & Chamot, A. (1990). Strategies used by Second Language Learners Cambridge: CUP.

Ott, L. D., Michailova, S., Silteipen, B., & Murchaika, V. (2016). Cultural intelligence: A review and new research avenues. *International Journal of Management Reviews*, 20(1), 99-119.

Padden, M. L. (2013). A pilot study to determine the validity and reliability of the level of reflection-on-action assessment. *Journal of Nursing Education*, 52(7), 410–415.

Palladino, P., Cornoldi, C., De Beni, R., & Pazzaglia, F. (2001). Working memory and updating processes in reading comprehension. *Memory and Cognition*, 29, 344–354.

Pieschl, S. (2009). Metacognitive calibration – an extended conceptualization and potential applications. *Metacognition and learning*, 4(1), 3-31.

Plakht, Y., Shiyovich, A., Nusbaum, L., & Raizer, H. (2012). The association of positive and negative feedback with clinical performance, self-evaluation and practice contribution of nursing students. tiny.cc/NET_feedback. Last accessed: June 14, 2017.

Rahimi, M. & Katal, M. (2010). Iranian university students' metacognitive listening strategies awareness in learning English. Paper presented at the third national conference on education, Tehran, Iran.

Rawson, K. A., O'Neil, R., & Dunlosky, J. (2011). Accurate monitoring leads to effective controls and greater learning of patient education materials. *Journal of Experimental Psychology: Applied,* 17, 288–302.

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General,* 137, 615–625.

Sart, G. (2014). The effects of the development of metacognition on project based learning. *Procedia - Social and Behavioral Sciences*, 152, 131-136.

Schellings, G. (2011). Applying learning strategy questionnaires: Problems and possibilities. *Metacognition and Learning*, 6, 91–109.

Serra, M. J., & DeMarree, K. G. (2016). Unskilled and unaware in the classroom: College students' desired grades predict their biased grade predictions. *Memory and Cognition*, 3, 1–11.

Siebig, S., Kuhls, S., Imhoff, M., & Gather, U. et al. (2010). Intensive care unit alarms – how many do we need? *Critical Care Medicine*, 38, 451–456.

Smith, D., & Dumont, F. (1997). Eliminating Overconfidence in Psychodiagnosis: Strategies for Training and Practice. *Clinical Psychology: Science and Practice*, 4 (1), 335-345.

Smith, K. (2005). Teacher educators' expertise: What do novice teachers and teacher educators say? *Teaching and Teacher Education*, 21, 177–192.

Soderstrom, N. C., & McCabe, D. P. (2011). The interplay between value and relatedness as bases for metacognitive monitoring and control: Evidence for agenda-based monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 1236–1242.

Stankov, L., & Crawford, J. (1997). Self-confidence and performance on tests of cognitive abilities. *Intelligence.* 25. 93-109.

Stone, N. J. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review*, 12 (4), 437-473.

Swanson, H. L. (1990). Influence of metacognitive knowledge and aptitude on problem solving. *Journal of Educational Psychology*, 82(2), 306-314.

Szpunar, K. K., Jing, H. G., & Schacter, D. L. (2014). Overcoming overconfidence in learning from video-recorded lectures: Implications of interpolated testing for online education. *Journal of Applied Research in Memory and Cognition*, 3, 161–164.

Tauber, S. K., & Rhodes, M. G. (2010). Does the amount of material to be remembered influence judgements of learning (JOLs)? *Memory,* 18, 351–362.

Thiede, K. W., Redford, J. S., Wiley, J., & Griffin, T. D. (2012). Elementary school experience with comprehension testing may influence metacomprehension accuracy among seventh and eighth graders. *Journal of Educational Psychology*, 104, 554–564.

Thomas, T., & Walsh, R. (2008). Multifaceted orientation to foster reflective thinking, metacognition and critical thinking in new graduate nurses. *Critical Care Nurse*, 28 (2), 13–19.

Thompson, V. A., Turner, J. A. P., & Pennycook, G. (2011). Intuition, reason and metacognition. *Cognitive Psychology*, 6(1), 107-140.

Tobias, S., & Everson, H. T. (2003). Knowing what you know and what you don't: Further research on metacognitive knowledge monitoring. College Board Research Report No. 2002-03. College Entrance Examination Board, New York.

Wahlheim, C. N. (2011). Predicting memory performance under conditions of proactive interference: Immediate and delayed judgements of learning. *Memory and Cognition*, 39, 827–838.

Weaver, R. M. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment and Evaluation in Higher Education*. 31(3), 379-394.

Winne, P. H., & Jamieson-Noel, D. L. (2002). Exploring students' calibration of self-reports about study tactics and achievement. *Contemporary Educational Psychology*, 27, 551–572.

Winne, P. H., & Nesbit, J. C. (2009). Supporting self-regulated learning with cognitive tools. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), The educational psychology series. Handbook of metacognition in education (pp. 259-277). New York: Routledge/Taylor & Francis Group.

Wood, B. P. (2000). Feedback: a key feature of medical training. *Radiology*, 215, 17–19.

Yaeger, K. A., & Arafeh, J. (2008). Making the move: From traditional neonatal education to simulation-based training. *Journal of Perinatal & Neonatal Nursing*, 22, 154–158.

Zabrucky, K. M., Agler, L. L., & Moore, D. (2009). Metacognition in Taiwan: Students' Calibration of Comprehension. *International Journal of Psychology*, 44 (4), 305-312.

CHAPTER III

DETERMINING THE IMPACT OF MODIFYING ALARM THRESHOLD LIMITS SET BY DEVICE MANUFACTURERS ON ALARM RESPONSE AND ERROR RATE

3.1 Abstract

Objective: This study aimed to investigate the effect of customizing physiological monitor alarm settings based on patient clinical conditions on alarm response rate and committed error rate. Further, the study also attempted to establish the relationship between different alarm settings and alarm response rate.

Methods: Thirty participants, nurses and their assistants, managed alarms on a physiological monitor under two conditions – default and modified – in clinical simulator. Cumulative alarm response rates were recorded for each condition, and error rates were computed based on the number of errors committed out of the total number of alarms addressed. Patient care experience and satisfaction level for both conditions were also collected via a survey.

Results: Participants addressed more alarms and committed less errors when alarm threshold limits were modified based on clinical conditions. Post experiment survey results revealed that customization of alarm limits increased care provider experience and overall satisfaction.

Conclusion: Results from this exploratory study show that customization of alarm threshold limits will improve nurses' alarm response rates and reduce the number of errors they commit while improving patient safety and outcome. Although this study's smaller sample size and controlled setting limits the generalizability of results, this one-equipment in one-unit pilot study can be considered as a starting point for reviewing alarm management processes across the board – multiple units, devices and conditions. This study clearly establishes that default alarm limits produce too many alarms to manage within a short period and that making minor adjustments to alarm setting will result in significant benefits.

3.2 Introduction

Nurses and nurses' assistants frequently rely on physiological monitors to watch patients in their care units. They rely on these monitors to alert them when a serious problem occurs. Alarms on these monitors are intended to alert them when deviations arise from a predetermined normal status. There are mainly two types of alarms – alarms associated with clinical conditions and alarms due to equipment condition. Since alarms indicating equipment or technical conditions do not require clinical intervention and are clearly differentiated by device manufacturers, they will not be considered in this study. Device manufacturers typically differentiate machine condition alarms through different frequencies, harmonics, and variations in tone. The focus of this study will be on alarms associated with clinical conditions. These alarms are further classified into three types - actionable alarms, non-actionable alarms and false alarms. An actionable alarm is any alarm due to a true underlying clinical condition that requires a clinical intervention or triaging with other nurses/assistants. A non-actionable alarm is any alarm that is valid for

a condition that is usually transient and does not require immediate intervention or triaging. A false alarm is defined as any alarm triggered due to incorrect identification of an underlying condition or due to interference with other systems or set up. Artifacts and low voltage asystole are examples of false alarms. A problem develops when the number of alarms to monitor increases and becomes overwhelming to nurses/assistants. Nurses/assistants providing care become desensitized to the alarms when the frequency of the alarms increases (Edworthy & Hellier, 2005).

Through systematic review and experimental studies, researchers have established that the cardiac monitor algorithms provided by device manufacturers are highly sensitive and are a key contributor for non-actionable and false alarms. These non-actionable and false alarms not only interfere with patient care, but also reduce trust in all alarms. Nurses/assistants often find methods to work around or overcome these alarms or start ignoring them completely. (Sanderson et al 2006). Research shows that non-actionable and false alarms occur in the range of 86% to 99.4% (Edworthy & Hellier, 2005). There is the potential for an actionable alarm getting ignored or missed in this myriad of non-actionable and false alarms.

3.3 Background

In the past five years, medical errors occurring within healthcare organizations have increased. The consequences of these errors range from minor to catastrophic for healthcare recipients (Edwards & Moczygemba, 2010; James, 2013). In order to monitor healthcare recipients who have been admitted to an intensive care unit (ICU), healthcare staff must consider large quantities of highly heterogeneous information, including past medical history, X-rays, ultrasound scans, laboratory analyses, and data from

examinations. Out of these items, the greatest overload of work results from the monitoring of physiological variables: electrocardiogram, heart rate, blood oxygen saturation, end-tidal CO_2 rate, blood pressure, etc. (Hannibal, 2011). These variables represent underlying patient conditions and may change over time. Changes to these variables often represent the appearance of physio-pathological processes that require rapid intervention to mitigate or avoid life-threatening situations for the patient. Thus, they are constantly mentally demanding and cause operator fatigue (Bell, 2010).

Alarm fatigue is frequently identified as a patient safety issue (Siebig et al., 2010). Alarm fatigue occurs when a caregiver becomes overwhelmed by a large number of clinical alarms such that important alarms can be missed or ignored (Burgess, 2009). The Joint Commission, the American Association of Critical-Care Nurses (AACN), the Food and Drug Administration (FDA), the ECRI Institute, and the Association for the Advancement of Medical Instrumentation (AAMI) have all identified the need to address alarm management and alarm fatigue (Sendelbach & Funk, 2013; The Joint Commission, 2013). In 2011, AAMI convened a Medical Device Alarms Summit, and after the summit, created an alarm best practices workgroup to address the problem of clinical alarm fatigue. In 2012, 2013, and 2014, the ECRI Institute (formerly known as Emergency Care Research Institute and later changed to abbreviated name of ECRI) identified clinical alarm hazards as the top potential danger area in hospitals and health systems (ECRI, 2014; ECRI, 2013; ECRI, 2012).

The aforementioned non-profit and government entities are working hard to determine ways to find a solution to this deeply complex problem. Improvements in technology, specifically to devices' sensitivity and specificity, have been a major focus of

the medical technology industry. A few researchers have demonstrated that introducing a small time-delay within the alarm algorithm can reduce the number of alarms by as much as 30%-46% and alarm fatigue by 30% -50% (Graham & Cvach, 2010). Alarms are triggered by many underlying events when machines are set at a conservative range (textbook normal values) and tend to disappear within that range when algorithms recognize the fluctuations (Barker, 2002). Therefore, introducing a time delay between the onset of the trigger and the onset of the alarm will reduce the number of false alarms. In an interventional study by Borowski et al. (2011), a time delay of 5 to 19 seconds was introduced and a 70% reduction in alarms was observed during the 200-hour study period without compromising the integrity and clinical safety of the system. In an alarm setting modification study by Welch (2011), a time delay of 5 seconds reduced the alarm frequency by as much 32%.

As many alarms are in SpO_2 (a measure of arterial oxygen saturation), monitoring situations are caused by self-correcting desaturations. Introducing minor delays would therefore be an easy way to reduce unnecessary alarms. It was also demonstrated that by reducing the SpO_2 rate by 2% from a default setting of 90%, false alarms could be reduced by as much as 45% (Drew et al., 2014; Gazarian, 2014). Many institutions use the default setting as a standard setting (Graham & Cvach, 2010; Nix, 2015; Sendelbach et al., 2015). Graham and Cvach proposes that further reduction in threshold limit should be implemented for additional benefits such as patient comfort, better ambient noise and improved clinical outcome. The study recommends reviewing patients' condition and resetting the alarm threshold limit in physiological monitors. A similar recommendation is made by Edworthy (2013); in his review, Edworthy estimates that a six-fold reduction

in alarm frequency can be achieved by reducing SpO_2 to 85% and introducing a time delay of 15 seconds. Based on these studies, one can see that there is a sizeable opportunity for reducing alarms by customizing alarm threshold limits.

3.4 Customization on patient level

It is well known that in the past three decades, medical devices have increased in complexity, and many of them are designed with alarm systems and notification mechanisms as part of a risk mitigation strategy. Currently available physiological monitoring systems provide alarms on most physiological data with high accuracy (Nix, 2015; Bonafide et al., 2015). Thus, these systems produce a great number of alarms and signals that must be managed. It is possible to count more than a dozen alarm sources in a typical step-down recovery unit, taking into account ventilation data, electrocardiogram, and arterial pressure and pulse oximetry for a patient (Billinghurst et al., 2003; Dandoy et al., 2015). Alarms generated by the infusion pump, the nutrition pump, the therapeutic control systems and the dialysis system, among others, should also be added to this list.

The present technique used to generate an audible alarm signal is based on setting a default threshold. There is no standard for default alarm settings. For a given parameter, this default setting can vary from one monitoring system to another. In some cases, the manufacturer recommendations are considered as default settings and in others, the last used settings are considered as default settings. In addition, requirements established by hospital policies and procedures vary significantly resulting in non-standardized limits for default setting (Varpio et al., 2012). One of registered nurses' fundamental roles is to identify signs and symptoms of deterioration in their patients' conditions and act to interrupt continued deterioration (Boev, 2012). Hospitals use patient monitoring

equipment to continuously monitor patients, so that any critical or life-threatening conditions can be detected and acted upon by care providers (Billinghurst et al., 2003). Several types of signals such as acoustic alarms, voice-overs, visual image transmission systems, and voice or text alert systems are used to communicate a multitude of conditions (Cvach, 2012; Konkani et al., 2012).

Due to technological advancements and a competitive market, device manufacturers have made tremendous progress in their devices' sensitivity to and accuracy of measured physiological responses (Taenzer et al., 2011). Although this progress is desired, these strides in technology do not translate well from development and simulated test environments at the manufacturers' sites to real-world clinical applications. Patient monitors, the devices that alert clinicians about changes in physical and chemical signals (broadly referred as physiological signals), and therapeutic and delivery devices, the devices that deliver fluids and therapy, are ubiquitous these days (Talley et al., 2011). The alarms that these devices generate are excessive and cause a phenomenon widely known as 'alarm fatigue'. This occurrence is generally characterized by a clinician's desensitization to alarms when the number and frequency of alarms that need to be monitored or addressed becomes overwhelming (Cvach, 2012; Funk et al., 2014).

The official definition of alarm fatigue, as drafted by the ECRI, refers to a sensory overload for staff who are exposed to an excessive number of alarms. Because of this overload, desensitization to alarms can occur and result in missed alarms. The consequence of alarm fatigue ranges from simple clinician dissatisfaction to poor quality patient care and compromised patient safety. Acknowledging the excessive alarms as a

problem, the Joint Commission named alarm fatigue the largest contributing factor to alarm related sentinel events in hospitals (The Joint Commission, 2015). The Joint Commission was not the only agency to recognize this as problem; other non-profit quasi-government organizations such as the Emergency Care Research Institute (ECRI) and the Association of Medical Instrumentation (AAMI) have identified alarm fatigue as a critical issue (Mitka, 2013). When nurses experience alarm related fatigue, they delay their responses; fail to respond, possibly disabling, or silencing alarms, which could compromise patient safety (Shekelle et al., 2015). Alarm fatigue may also cause nurses to resort to unsafe workarounds such as reducing the volume, pausing alarms or disabling them altogether (Shekelle et al., 2011). Even after gaining multiple watchdog agencies' attention, little empirical data has been collected on alarm fatigue by hospitals (Sowan et al., 2015). The little research conducted by researchers on alarm fatigue is limited to alarm reduction techniques, improved algorithm, alarm response time, and alarm fatigue's impact on patient care quality. In order to find an appropriate solution, we need to understand the magnitude of the problem. Previously conducted studies were all regarding intensive care settings or specialized and focused care settings. As many hospitals have tertiary care centers (up to 35% of hospitals according to the American Hospital Association; Fast facts 2017), it is important to understand issues related to alarm fatigue in that setting. This study will help us understand the relationship between alarm response rate, error rate and care provider satisfaction in tertiary care settings.

3.4.1 Various alarms

It is estimated that 85%–99% of alarms do not require an intervention (Cvach 2012). Sowen et al. (2016) state, "the problematic high volume of false and clinically

insignificant nonactionable true positive alarms – up to 99.4% – results in clinician's failure to appropriately respond." Causes for this high percentage of alarms not requiring clinical intervention include setting the alarm thresholds 'too tight,' not adjusting default alarms to individual patient needs, or incorrectly applying sensors. Research has documented that a significant proportion of patients placed on ECG-telemetry do not meet the American Heart Association indications for telemetry monitoring and are not deemed to be at increased risk for irregular heart rhythm (Billinghurst et al., 2003).

Clinicians rely on the information from signals, alerts and alarms generated by medical devices to understand their patient's current state of well-being and how it changes over time (Tsien & Fackler, 1997; Funk et al., 1997). Therefore, it is imperative to reduce or remove the hazard associated with an excessive number of alarms and only allow required alarms to gain nurses' attention so that they can function better and provide the best possible care (Bitan et al., 2004). Early recognition of deterioration through vital signs can logically be assumed to prevent adverse events such as delay in diagnosis and treatment. To address alarm fatigue appropriately and adequately, one should understand the various types of alarms that exist in a typical hospital setting. The ACCE Healthcare Technology Foundation classifies alarms into three broad categories (ACCE white paper on alarms, 2006):

3.4.1.1 Actionable alarm

An actionable alarm is an alarm that requires a clinician's intervention or warrants a clinician's input or interaction with other clinicians or patients. This alarm should lead to immediate intervention, but due to alarm fatigue could go unwitnessed or misinterpreted by the attending clinician. Actionable alarms require timely intervention to prevent an adverse event.

3.4.1.2 Non-actionable alarm

This alarm correctly identifies the underlying patient's physiologic condition, but does not require intervention. Its validity is based on waveform quality and accuracy, strength of signals from leads and detectors, and artifact conditions. Transient low-oxygen saturation, non-critical arrhythmia and heart rate alarms are a few examples of non-actionable alarms (Manzey at al, 2014). In a majority of cases, these short duration alarms correct themselves. Some of these alarms may require contextual information to understand better. Repetition of many non-actionable alarms may be a precursor to a true, valid actionable alarm, but audible tone does not necessarily require a response every time it occurs (Getty et al, 1995). Non-actionable alarms that capture clinicians' attention but are not clinically significant contribute heavily to alarm fatigue.

3.4.1.3 False alarm

Alarms caused by patient motion, poor sensor placement, bent pins, connection error, cable issues and limitations in the device alarm detection algorithm are referred to as false alarms(Chambrin et al, 1999). These alarms could also be generated by bad or missing data. The majority of alarms generated due to equipment condition or technical

condition are non-actionable alarms. The classification of alarms and at what point intervention is required from nurses and their assistants is shown in Figure 3.1.

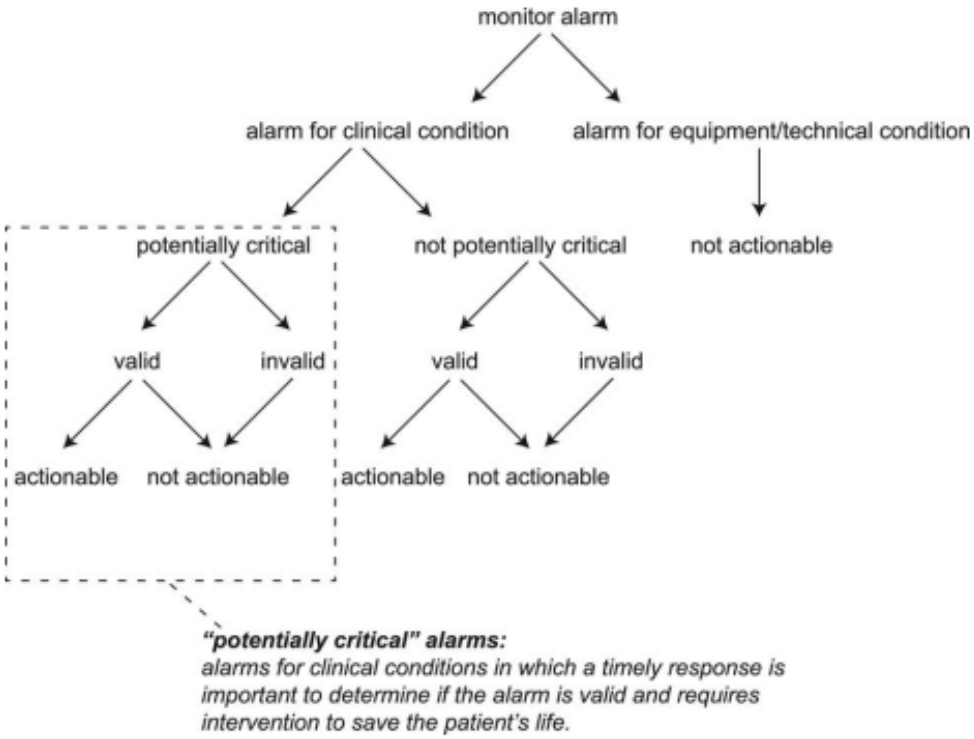

Figure 3.1 Categories of alarms

Adapted from Association Between Exposure to Nonactionable Physiological Monitor Alarms and Response Time in a Children's, by Bonafide et al., 2015, Hospital Journal of Hospital Medicine, 10 (6), pp 345.

As patient safety relies upon distinguishable alarms by a team of cross-functional healthcare professionals, the International Electro-technical Committee (IEC) attempted to standardize alarm types and tones through an International Standard IEC 60601-1-8. This standard includes 17 melodic alarms. Research indicates that listeners can only learn to recognize 4 to 6 alarms (Gazarian, 2014); this number only increases up to 12 after weeks of practice (Chambrin et al., 1999; Mitka, 2013). A study on the

discriminability of alarms conducted by Sanderson et al. (2006) shows that participants' abilities to recognize alarms and the accuracy of tasks completed dropped when the workload (from alarms) was increased. Therefore, it is important to discern the few actionable alarms from the overwhelming amount of alarms and adequately respond to them without any errors when providing patient care. In addition to alarm limit modification, a hospital can take a number of other practical measures to lower false alarm rates, such as a proper maintenance program for monitors, standard protocol for electrode placement and policies for skin preparation (Barker, 2002;Atzema, et al., 2006; Cvach, 2012). Since it is easy and inexpensive, alarm rate modification should be the first step in reducing the number of alarms that need to be addressed. Findings from this study emphasize the importance of individualizing alarm limits in tertiary care settings.

3.4.2 Alarm threshold limit and its impact

Since the publication by Lawless on the "crying wolf" phenomenon in 1994, false alarm rates have remained stubbornly high and there has been an unresolved concern in hospitals (Lawless, 1994). These false alarms can lead to disrupted care, affecting both the patient and their care providers through noise disturbances and slower response times (Bliss et al, 1995). False and non-actionable alarms as high as 86% have been reported by previous studies in alarm management (Whalen et al., 2014; Gazarian, 2014). Cvach et al. (2012) have reported that there can be as many as 350 alarms/patient/day at a typical hospital in the United States, which equates to 2.5 million alarms in a year in a ward of 20 beds. Even if all of the false alarms were successfully eliminated, it would leave up to 140,000 alarms that need to be answered. Therefore, there is a pressing need to learn how

nurses respond to alarms, determine the error rate and compute the workload index under different alarm conditions, so that better solutions can be prescribed.

A pulse oximeter is the most commonly used medical device in a hospital setting (Atzema et al., 2006). The pulse oximeter noninvasively and painlessly measures peripheral arterial oxygen saturation (SpO_2) and pulse rate with a sensor that is placed on upper or lower extremities. Fluctuating oxygenation range below 90% reflects a clinically significant event known as a hypoxemic episode (Drew et al., 2014). Pulse oximeter alarms are triggered to alert nurses of a possible hypoxemic episode when SpO_2 falls below a threshold for a pre-specified period of time (alarm delay time). These alarms can also be triggered by non-actionable or non-clinically relevant events such as patient motion or attributed to a unique patient condition (such as massive pulmonary embolism). Due to their ubiquitous nature, pulse oximeters are one of the highest alarm generators in hospitals (Funk et al., 2014). Hospitals' biomedical engineering departments typically set new pulse oximeters at their default setting or at manufacturer recommendations of 90% SpO_2. However, research shows that only 30% of admitted patients meet this normal threshold, and the remaining patients tend to have a value lower than this limit due to various factors and pathophysiological conditions (Chambrin et al., 1999, Otero et al., 2009; Manzey et al., 2014). Biomedical engineering departments can reduce false alarms in two ways -- by altering the threshold based on the unit (intensive care, step-down unit, progressive unit), which will then apply to every patient entering the unit, or based on the individual patient. Similar to physiological monitors, numerous other medical devices generate alarms in the patient care environment that can also contribute to alarm fatigue and which need to be evaluated thoroughly.

Graham and Cvach (2012) removed duplicate alarms in their study and noticed a significant reduction in true valid alarms. Duplicate alarms are alarms and notifications or alerts for the same underlying condition. Graham and Cvach note that the alarms for high and low heart rate were the same as the bradycardia and tachycardia alarms. The algorithm for computing bradycardia was slightly different from the low heart rate calculation, and the same applies for tachycardia. The most common cause of false (asystole) alarms is under-counting of heart rate due to failure of the device to detect low-voltage complexes in the ECG leads used for monitoring (Blum et al., 2010). The parameters for tachycardia and higher heart rate may be mathematically different and can reflect the difference in algorithm and formulas; however, they may not make much difference clinically. Therefore, Blum et al changed lower and higher rate alarms to message level (without audible tone) and increased the alarm tone for bradycardia and tachycardia to a warning level. Surveyed nurses' patient care satisfaction increased significantly after this simple change, and nurses reported reduced alarm associated fatigue (Graham & Cvach, 2010).

Due to liability lawsuits and for competitive reasons, device manufacturers design devices with maximum sensitivity -- alarms are devised to emit audible tones for every true condition, but the devices could also trigger alarms for every tracing or minor low-voltage deviation in the sensitivity algorithm (Manzey et al, 2014). Researchers recommend Machine Learning concepts to be introduced within the algorithm -- smart contextual pattern sensing and self-reprogramming of algorithms (Otero et al., 2009). For example, in watching a patient with an atrial fibrillation condition, a smart algorithm will trigger alarms when there is a significant percentage (as set by the Physician) change in

heart rhythm, rather than repeatedly triggering an alarm for every fibrillation change (Pelter and Drew, 2015). Pelter and Drew further recommend integrating alarm algorithms with physiological parameters such heart rate and blood pressure. These changes will help to identify if alarms are real or false in the context of physiological data. For example, an asystole event unaccompanied by a change in blood pressure would be re-classified as a system message or false alarm, and hence would not trigger an alarm (Sowan et al., 2016).

In a study conducted by Sowan et al. (2015), 38% of those surveyed reported that they do not change alarm parameters at all, and 20% of nurses reported that they modify alarm threshold only if needed. Nurses have reported a lack of confidence in customizing default settings to be patient specific, recognizing when specific monitoring is needed for specific medical cases, and in eliminating redundant alarms when changing default settings. Increasing unit nurses' awareness of patient conditions and having a closer contact with dispatching physicians during the hand-off may be useful in tailoring alarm settings and improving alarm management (Oliver et al., 2014). Manipulation of monitor defaults and staff training are not sufficient to sustain change unless the unit is held accountable for maintaining a zero tolerance for nuisance alarms and troubleshooting these alarms as soon as they occur.

Data collected by Sowan et al. (2016) found that the complexity in navigation to set alarm parameters, among a few other factors, contribute to a high percentage of nurses not attempting to modify alarm settings. The price we pay for not taking a simple step of modifying alarm thresholds is very steep, ranging from annoyance to death (Korneiwicz et al., 2008). Therefore, researchers recommend a multi-pronged alarm management

approach that includes several factors such as various alarms in the organization, unit nurses' understanding of alarms, training, team dynamics, normal alarm response time and safety culture (Sowan et al., 2015). While previous studies have evaluated response time, they have not evaluated response rate, committed error rate, and their relationship to care provider experience satisfaction during alarm management. In this study, we evaluated these parameters and their relationship to overall care provider satisfaction under two different settings (default alarm setting and modified alarm setting). Results from this study contribute to our understanding of alarm fatigue in a step-down unit under tertiary settings. Furthermore, quantification of cognitive workload load index during alarm management will help in identifying sources of stressors and will direct valuable resources in addressing the root-cause of the problem.

3.5 Hypotheses

Specific hypotheses investigated include:

1. Alarm response rate will be significantly higher when alarm threshold limits are modified. In other words, there will be a difference in response rate for modified alarm threshold limits. [H_0 = No difference]

2. Nurses/assistants will commit fewer errors when the settings are modified. In other words, there will be a difference in committed error rate for modified alarm threshold limits. [H_0 = No difference]

3. Nurses'/assistants' patient care experience and overall satisfaction will be higher when physiological monitor settings are modified.

3.6 Methods

In this section, the experimental methods and participants are defined to examine the impact of altering alarm threshold limits, according to individual patient's clinical condition, on alarm response and error rates. Hypotheses from the previous section are incorporated as experimental variables.

3.6.1 Experimental design

A between-subjects ANOVA and Wilcoxon-Mann-Whitney U median rank test were used to test for differences in alarm response and committed error rates between two alarm conditions (normal-default and modified threshold settings) in a simulated progressive patient care setting. Response rate was calculated by the number of alarms attended for a given number of alarms presented. Error rate was computed by the number of incorrectly addressed alarms for a total number of attended alarms. Patient care experience and satisfaction level data for each alarm setting were collected from participants and assessed through a post-experiment survey.

3.7 Variables

The independent variable tested in this study was alarm threshold settings. The two levels of alarm threshold settings included were: (i) default (as set by the manufacturer) and (ii) modified (for a simulated patient condition). All participants, while addressing alarms, completed normal patient care tasks that are typical in a progressive patient care setting to closely mimic a real-life situation.

Four dependent variables were measured in this study: (i) alarm response rate (measured in terms of alarms responded out of the total number of alarms presented), (ii)

error rate as % (accuracy or correctness of responded alarms). Example: If nurses/assistants accurately responded to 13 alarms out of 16 presented alarms, the error rate will be 19%, (iii) care provider experience, and (iv) overall satisfaction.

3.8 Participants

Participants for this study included 30 nurses (these were different participants from previous studies) from various local area hospitals. Refer to the next paragraph for sample size determination. Because using the same participants from previous studies/experiments may alert the participants to the intent of the study and tempt them to focus on alarm related tasks, a new set of participants were used in the study. The sample population composed of 23 females and 7 males, ranging from 24 to 60 years of age (M = 40.66 years, SD = 9.85 years). Flyers and word of mouth were used to recruit participants. Recruited participants were randomly assigned to one of the two alarm threshold groups – default alarm setting and modified setting. Inclusion criteria for the study included medical alarm exposure and basic patient care experience. There were no exclusion criteria for this study.

Previous studies were examined to determine the appropriate sample size. A study by Graham and Cvach (2010) utilized 30 nurse participants in examining the impact of modified alarm limits on fatigue associated with alarm management. Sample sizes for previous studies in assessing alarm response time ranged from N = 26 (Bonafide, et al., 2015) to N= 9 (Gazarian, 2014) in a study which included frequency and type of alarm. Sample size was set at N = 48 for a pilot study on alarm settings for critically ill patients by Christensen et al. (2014). Based on this review, a sample size of 30 was used for this study.

3.9 Experimental protocol

This study assessed the effect of modifying alarm limits on participants' response and error rates while providing patient care in a simulated setting. The entire experiment was conducted in two waves over the course of two weeks. A week was dedicated for each alarm setting – default alarm threshold and modified setting. The patient condition to be monitored was kept constant to reduce variability. As each hospital/unit sets alarm management protocol, local area hospitals in the Pacific Northwest were referred and the protocol followed at a 412-licensed bed hospital's progressive care unit was used for this study. The default setting alarms shown in Table 3.1 is based on standard protocol. Since previous studies show a typical nurse in a progressive care unit does not spend his/her entire time solely on alarm management and performs other duties for three patients (Spence & Leiter, 2006; Falk & Wallin, 2016; Clark & Yoder-Wise, 2015), a similar set up was reproduced in a clinical simulator for this experiment. The details of simulated tasks and flow are provided in subsequent sections. The other tasks performed by nurse participants are called dummy tasks and not included in data analysis. The details of dummy tasks are provided in section 3.9.3. Participants were strongly encouraged to complete all dummy tasks. These dummy tasks were also kept at the same difficulty level between different alarm conditions (normal alarm threshold and modified setting) to minimize variability. The randomly assigned participants were verbally briefed at a high level (elaborate details about alarm management, response and error rate were not revealed to limit participants' bias toward alarms) about the experiment prior to starting the experiment, and were offered an opportunity to ask any questions or to ask for clarifications. During the briefing session, details about tasks to be performed,

mannequins' conditions, total experiment time, various alarms and dummy tasks were provided.

3.9.1 Setup

The simulator setup for experiments was a progressive step-down care unit (patients moved to this unit are typically low risk and in the recovery phase for their clinical condition). Three male patient mannequins (SimMan®), identified as M-1, M-2 and M-3, were placed in supine positions and identified as low risk based on the Goldman risk chart. M-1 was instrumented with a ProSim SpotLight® pulse oximeter simulator (Fluke Bio, Bothell, WA). A physiological monitor (Nellcor® with software algorithm "Smart SatSec®" feature for customization) connected to the pulse oximeter simulator presented alarms shown in Table 3.1. The physiological monitor was set at default level for the default-setting portion of the experiment, and the Smart SatSec® was utilized for the modified setting. Alarms, shown in Table 3.1, were presented on the screen at a programmed time interval using auto sequence mode. For both settings, the software algorithm was programmed to keep the alarm available for 75 seconds and automatically stop when the time lapsed. The alarm sequence, type and characteristics are discussed in section 3.9.2. M-2 and M-3 were not required to be monitored; they were simply recovering from minor outpatient surgical procedures. These mannequins were part of the experiment to emulate a progressive care unit as close as possible. Nurse participants performed other assigned tasks on these mannequins (M-2 and M-3) as part of the experiment. The whole session was observed through a one-way mirror in the simulator, and the experimental data was recorded.

3.9.2 Default parameters, various alarms and sequence

Default alarm parameters at a progressive step-down care unit using Nellcor™ PM 1000N pulse oximetry system and the modified alarm setting utilizing "Smart SatSec™" for a patient mannequin is shown in Table 3.1. There is approximately 38% reduction in total number of alarms to manage when the alarm thresholds were modified.

[Intentionally left blank]

Table 3.1 Various alarms sequence

S.No	Default Setting (as released to the hospital floor)		S.No	Modified to patient condition using "Smart SatSec™"	
1	Advisory Alarm	Non-Actionable	1	**Removed[1]**	
2	**Warning Alarm**	**Actionable Alarm**	2	**Warning Alarm**	**Actionable Alarm**
3	System Message	Non-Actionable	3	**Removed[1]**	
4	**Actionable Alarm**	**Actionable Alarm**	4	**Actionable Alarm**	**Actionable Alarm**
5	**Warning Alarm**	**Actionable Alarm**	5	**Warning Alarm**	**Actionable Alarm**
6	System Message	Non-Actionable	6	**Removed[1]**	
7	**Warning Alarm**	**Actionable Alarm**	7	**Warning Alarm**	**Actionable Alarm**
8	**Actionable Alarm**	**Actionable Alarm**	8	**Actionable Alarm**	**Actionable Alarm**
9	**Warning Alarm**	**Actionable Alarm**	9	**Warning Alarm**	**Actionable Alarm**
10	System Message	Non-Actionable	10	System Message	Non-Actionable
11	System Message	Non-Actionable	11	**Removed[1]**	
12	Advisory Alarm	Non-Actionable	12	**Removed[1]**	
13	**Warning Alarm**	**Actionable Alarm**	13	**Warning Alarm**	**Actionable Alarm**
14	Advisory Alarm	Non-Actionable	14	Advisory Alarm	Non-Actionable
15	**Actionable Alarm**	**Actionable Alarm**	15	**Actionable Alarm**	**Actionable Alarm**
16	System Message	Non-Actionable	16	System Message	Non-Actionable
17	Advisory Alarm	Non-Actionable	17	**Removed[1]**	
18	Advisory Alarm	Non-Actionable	18	**Removed[1]**	
	Total no of alarms = 18			**Total number of alarms = 11**	

[1]Removed alarms were: 5 PVCs (premature ventricular contraction), 1 missed beat, and 1 noninvasive blood pressure.

3.9.3 Task details

Nurse participants (one at a time) were asked to check in at one of the nurse bay stations in the simulator. Before starting the session, participants were provided with an overview of the experiment, and asked to complete an informed consent form approved by the Mississippi State University IRB (Appendix A) and complete a paper demographic questionnaire (Appendix A). The entire experimental session lasted approximately 50 minutes and each nurse participant was presented with following the tasks (in the same sequence). During this session, the medical mannequin "M-1" presented alarms to study participants and the experimental data was recorded. Completion rates of tasks presented in this section were recorded but were not be analyzed statistically. The researcher reminded participants through the microphone when the task was due for completion. To minimize order and interference effects, a 15-minute 'warmup' period prior to starting the session and a 2-minute 'cooling' period between tasks were provided to participants. During the warmup period, the experimenter discussed alarms and scenarios and asked them to verbally explain their response. As interference effects between tasks may impact participants' alarm management, the tasks (tasks 1-4) were presented with a two-minute cooling period before and after. The experimenter used a timer to align the tasks within alarm management.

- *Task 1:* Call Pharmacy and check for the status of ordered medicine for patient Mannequin # 2 [Timing: 2 minutes into the experiment; call duration: 30 seconds]
- *Task 2:* Enter blood work result in Epic hospital system software for patient Mannequin # 3 [Timing: 10 minutes into the experiment; task duration: 2 minutes]
- *Task 3:* Administer a bolus dose of pain medicine for patient Mannequin # 2 [Timing: 14 minutes into the experiment; task duration: 1 minute]

- *Task 4:* Take a call from another hospital unit to receive a patient into this unit. [Timing: 19 minutes into the experiment; task duration: 2 minutes]

The calls were made through an intercom system from outside the simulator and participants were prompted using the simulator voice communication system at the appropriate time for calls to be made by them. They were provided with typed scripts for the calls they were to make, and typed scripts are included within the IRB packet. The entire script is provided below.

Researcher (via voice communication system): 'Participant Name, please make the call to pharmacy and check the status of Thyroxine'.

- Task 1: Participant (through the phone): *Hi, this is Participant Name, calling from intensive care unit floor # 3. I am the care provider for patient Mannequin #2. His date of birth is xx/xx/xxxx. His last name is spelled as a,b,c,d,e,f,g,h and first name is spelled x,y,z,h,j,k, l. I like to check the status of Thyroxine compound. The quantity ordered by treating Physician is 20ml. Treating Physician is Dr. John Doe.*
  Researcher (through the phone): *The order is complete. It will be delivered to you in 20 minutes. Who should we deliver this to?*
  Participant (through the phone): *Please deliver to me or to the floor Charge Nurse. She is our supervisor for today. Thanks.* **Task Complete**

- Task 2: Researcher (via voice communication system): *Participant Name, please enter the 3 lines of blood work result from the sheet provided on the laptop. Laptop is on your right hand side and at the corner of the room.* **Task Complete**

- Task 3: Researcher (via voice communication system): *Participant Name, please bolus the patient, Mannequin #2. Prefilled syringes are in the top cupboard. Please use the 5ml size.* **Task Complete**

- Task 4: Researcher (through the phone): '*Hi, this is XXXX YYYY; floor Supervisor at the West Block Critical Care Unit. We have the treating Physician orders to discharge Patient Name John Doe to your unit at 6 PM today. Patient record is 2016-014567 and she is a 63-year-old female. We have completed the green and yellow discharge sheets and informed the care provider for the patient. The patient has a mild edema on lower right leg and is asymptomatic. The patient is on low dose heparin and two vascular access devices one on each hand on the lower cubital vein. The patient is in stable condition. Can we go over additional details?*

Participant (through the phone): *Hi, this is AAAAA BBBBB at the step-down unit. Yes we can do the hand-off now but my shift ends in 5.30 PM so I will not be able to receive the patient. If you prefer to wait about an hour, you can speak with the second shift nurse. Otherwise, please call my supervisor Janet Doe at extension x3568 to find out how we can handle the situation.* **Task Complete**

3.10 Statistical analysis

Appropriate descriptive statistics were computed for all dependent variables and demographics data such as age, and clinical experience. As a new set of participants were used for this study, demographic data were collected and reported in Table 3.2. To determine whether there are any statistically significant differences between the mean alarm response and error rates, two one-way ANOVAs (Analysis of Variance) were used. Assumptions of the ANOVA model were tested using the Ryan-Joiner method at a significance level of 0.05. As the normality assumptions found to be violated, the Welch-ANOVA method was utilized to test hypotheses. A series of comparison tests of $\chi 2$ were performed to examine if subscales' scores differed as a function of demographic characteristics (*i.e.*, age, gender, years of experience as a nurse, and alarm management experience). No differences were noted across analyses ($p > .05$). A Wilcoxon median rank within subject was used to test for any differences in care provider experience and satisfaction levels for participants when managing alarms under two different settings. All results were considered significant at an alpha level of 0.05. The IBM SPSS statistical software package version 25 for Windows was used for all statistical analysis.

[Intentionally left blank]

Table 3.2 Demographics data

Variables	Mean (SD) or %
Age	40.6 (9.9) yrs
Gender	
Female	76.7
Male	23.3
Nursing background	
Registered Nurse	33.3
Nurse assistants (CNAs)	66.7
Years of experience in managing device alarms	
None	0.0
Less than 1 year	3.3
1-3 years	10.0
3-5 years	30.0
More than 5 years	56.7
Training on medical device alarms?	
Yes	33.3
No	66.7
Training provided by your institution adequate?*	
Yes	16.7
No	46.7
Did your assigned unit provide any training?*	
Yes	23.3
No	26.6
Educational background[1]	
CNAs/Other	66.7
Associates	13.3
Bachelors	13.3
Graduate and more	6.7
Any other certifications?*	
Yes	16.7
No	20.0

*– Percentage does not equal 100 due to missing responses; [1] – not equal to 100% due to rounding.

3.11 Results

Descriptive statistics for each of the dependent variables are provided in Table 3.3. As expected, the results were considerably different between the two alarm threshold settings. Normality of datasets was assessed using Ryan-Joiner method. Alarm response rates and error rates for both alarm settings (default and modified) were found to be normal. Modification of alarm threshold limits according to patient conditions allowed nurses to address higher number of alarms with higher accuracy. Correspondingly, care provider experience and satisfaction levels were also much higher for modified setting than default setting. Participants' ratings about the number of alarms that occurred while caring for patients is shown in Figure 3.2. About 70% of the participants, under modified condition, felt that the number of alarms presented were the right amount, and 50% of the participants in default setting indicated that there were too many alarms to manage.

Table 3.3 Descriptive statistics for dependent variables

Alarm Setting	Variable	Mean (SD)	Total
Default	% of alarms addressed	68.9 (10.5)	30
	Error rate	9.5 (6.0)	30
	Care provider experience[1]	2.6 (1.3)	30
	Overall satisfaction[1]	2.5 (0.9)	30
Modified	% of alarms addressed	86.7 (7.6)	30
	Error rate	2.6 (4.5)	30
	Care provider experience[1]	3.8 (0.8)	30
	Overall satisfaction[1]	4.3 (0.6)	30

[1] – Measured on 5-point Likert scale of 1-5 with 1 being very dissatisfied and 5 being very satisfied.

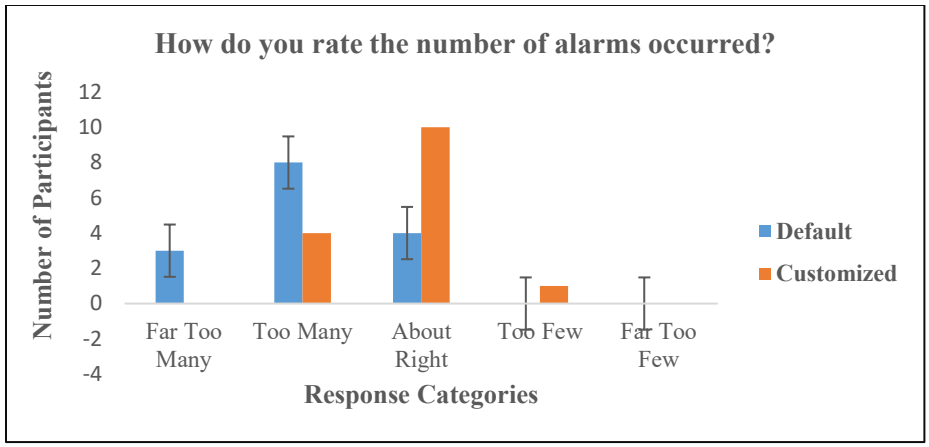

Figure 3.2 Participants response about number of alarms

3.11.1 Alarm response rate

A one-way Welch ANOVA was conducted to determine if the alarm response rate is different for different alarm threshold settings – default and modified. Participants were classified into two groups: default (n=15) setting and modified (n=15) setting. Alarm response rate was statistically significantly different between different alarm settings, Welch's F (1, 25.44) = 29.05, $p < .05$. Alarm response rate (i.e. number of alarms addressed) increased from the default setting to the modified setting due to fewer alarms when physiological monitoring is modified to patient conditions. A post-hoc analysis could not be conducted, as there were only two groups.

Table 3.4 ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2394.133	1.000	2394.133	29.053	.000
Within Groups	2307.333	28.000	82.405		
Total	4701.467	29.000			

Table 3.5 Welch's 'F' statistic

	Statistic[a]	df1	df2	Sig.
Welch	29.053	1.000	25.449	.000

[a]-Asymptotically distributed

3.11.2 Error rate

The error rate is defined as the total number of incorrectly addressed alarms out of the total number of alarms addressed during the experiment. For example, if a participant addressed 12 alarms, out of which 1 is incorrect, there is an error rate of 8%. A one-way Welch ANOVA was conducted to determine if the error rate is different for different alarm threshold settings – default and modified. Error rate was statistically significantly different between different alarm settings, Welch's $F(1, 25.93) = 12.46$, $p < .05$. ANOVA and Welch's test are shown in Tables 3.6 and 3.7, respectively. Error rate significantly decreased from default setting to modified setting, primarily due to fewer alarms when physiological monitoring is modified to patient conditions.

Table 3.6 ANOVA for committed error rate

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	346.800	1.000	346.800	12.462	.001
Within Groups	779.200	28.000	27.829		
Total	1126.000	29.000			

Table 3.7 Welch 'F' statistic

	Statistic[a]	df1	df2	Sig.
Welch	12.462	1.000	25.933	.002

[a]- Aysmptotically distributed

3.11.3 Care provider experience

A Mann-Whitney U test was run to determine if there were differences in care provider experience between default and modified alarm settings. Distributions of care provider ratings for default and modified settings were similar, as assessed by visual inspection. Care provider experience ratings (on a 5-point Likert scale) for modified setting (mean rank = 20.83) were statistically significantly higher than for default setting (mean rank = 10.17), $U = 32.5$, $z = -3.422$, $p = .001$, using an exact sampling distribution for U.

3.11.4 Overall satisfaction

To determine if there was any difference in overall satisfaction between default and modified alarm settings, a Mann-Whitney U test was run. Distributions of overall satisfaction ratings for default and modified settings were similar, as assessed by visual inspection. Overall satisfaction ratings (on a 5-point Likert scale) for modified setting (mean rank = 21.90) were statistically significantly higher than for default setting (mean rank = 9.10), $U = 16.5$, $z = -4.146$, $p = .001$, using an exact sampling distribution for U.

3.11.5 Regression

Regression was used to develop a regression model to predict the accuracy of the dependent variable alarm response rate. Stepwise model building was used to develop the model, with significant levels of entry and exit set to 0.05. The resultant model included only the default setting, as provided in equation 3.1

$$\text{Response Rate} = 86.80 - 17.87 \text{ Alarm Set\_Default} \quad (4.1)$$

This model indicates that alarm response rate drops by approximately 18% when default alarm setting changes by 1 magnitude, which means any increase, even smaller ones, will substantially reduce the alarm response rate. The ANOVA for the model is shown in Table 3.8. The $R^2 = 72.19\%$ indicating good model adequacy and total variation in response rate was explainable by alarm settings.

Table 3.8 ANOVA for regression model

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Regression	1.000	3394.000	72.190%	3394.000	3394.370	72.730	0.000
Alarm Setting	1.000	3394.000	72.190%	3394.000	3394.370	72.730	0.000
Error	28.000	1307.000	27.810%	1307.000	46.670		
Total	29.000	4701.000	100.000%				

3.12 Discussion

Hypothesis one stated that alarm response rate will be significantly higher when alarm threshold limits are modified. This hypothesis was supported by the results. Removal of alarms related to premature ventricular contraction, missed beat and noninvasive blood pressure allowed participants to respond better to remaining alarms and complete all other assigned patient care tasks. Majority of the patients checking into a progressive care unit do not need to be monitored for ventricular pacing or for missed

beats. These cases will result in frequent false nonactionable alarms and may strain caregivers. Similarly, older aged patients are likely to trigger higher number of ventricular contractions and missed beats alarms.

When alarm settings are modified, caregivers have fewer alarms to manage and thus have more time to provide quality care for patients and complete other essential duties. Removal of non-essential alarms alone is not customization; the effort could range from converting benign alarms to display messages and downgrading certain warning alarms to advisory notices for certain patients or delaying some alarms until they meet multiple criteria. All participants in modified alarm setting completed all assigned administrative patient care tasks whereas only 73% of the default setting participants completed assigned administrative patient care tasks. That is, 11 participants completed assigned 'other' administrate patient care tasks (tasks 1-4) and the remaining participants were not able to complete either task 2 or 3 after the voice-over prompt – they continued to resolve the alarm situation and skipped the 'other' task. This situation was considered as 'incomplete'. They likely received the voice-over and mentally processed it but could not complete as they spent their time in managing (excessive) alarms in default setting.

Based on these findings, one can conclude that any decrease in false or redundant alarms should result in a marked reduction in alarm burden with a higher proportion of clinically relevant alarms.

Hypothesis two stated that nurses/assistants will commit fewer errors when the settings are modified. This hypothesis was supported by the results from the Welch's ANOVA model. In other words, participants committed fewer errors when addressing alarms under modified setting. A likely reason for this outcome is fewer opportunities for

committing errors in modified settings (i.e. a manageable number of alarms would allow nurses to address the remaining relevant alarms appropriately). The alarm response rate is inversely proportional and the error rate is directly proportional to the total number of alarms that occurred, and these two dependent variables are closely related.

Hypothesis three stated that customization of physiological monitor settings would result in higher satisfaction levels and patient care experience. This hypothesis was supported by the results from the Mann-Whitney U test model. In the post-experiment survey, 5 out of 15 participants, (33%) in default setting, provided comments. Issues reported were focused on (1) the frequent alarms and (2) stress while managing the alarms and taking care of patients simultaneously. Although the same number of participants provided comments, the tone of those comments were strikingly opposite. Comments were around (i) ease of working the setting and (ii) calmness in the unit. Participants in the default setting were exposed to an excessive number of alarms that likely overloaded their senses, which led to frustration and limited cognitive readiness to attend other tasks. Based on these findings, we can see that participants were more consistent with their approach to alarm management in modified setting because the alarms that did go off had a higher probability of being a true alarm. The removal of non-actionable alarms in the progressive care eco-system has contributed to an environment where nurses were more attuned to the remaining alarms and addressed them more accurately and at a better rate. Minimizing alarms that are not actionable enhanced the environment of care, which in turn improved overall satisfaction for participants (Manzey et al, 2014). The removed alarms in modified settings were based on a complex interplay of incorrect user settings, underlying patient conditions, and algorithm deficiencies.

These factors could be rectified through a joint effort between nurses, hospital biomedical department, and device manufacturers without any major changes to hospital policies and procedures.

3.13 Conclusions

This study investigated the effect of customizing physiological monitor alarms on response rate, error rate, care provider experience and overall satisfaction while caring for patients in a progressive care setting. The results of the study clearly demonstrate that customization positively affects the number of alarms accurately addressed, care provider experience and overall satisfaction. The findings support the removal of non-essential alarms based on patient conditions. When these non-essential alarms, which contribute to sentinel adverse events and alarm fatigue, are removed, care providers will address remaining alarms accurately and have better job satisfaction. Though many organizations come together to provide care for a patient, nurses and their assistants ultimately bear the responsibility of managing and administering quality care delivered to patients. As such, nurses who work in a hospital setting can be exposed to considerable work-related stress, which typically results in burnout and reduced job satisfaction. Since a significant portion of nurses' and their assistants' work-lives are to diagnose and intervene when patients' clinical conditions change, which is frequently detected through medical devices and their alarms, it is important that manufacturers provide reliable device alarms and hospitals establish appropriate protocols and standard procedures.

The findings from this study are a small step in the right direction for hospital administrators and nurse managers who are involved in developing hospital policies and procedures for medical device alarms. Every physiological characteristic for every patient

do not need to be monitored – selecting a few that matter and omitting those that do not, will add value to patient care and will immensely benefit care providers and patients alike. Though the results are encouraging and tempt us to modify device alarm settings across the board, the study has many limitations. The results cannot be generalized without additional work.

3.14 Limitations and future research

The population sample of the participants needs to be expanded to additional populations such as physicians, medical assistants, and other therapists who are also part of the patient care team. The sample population was entirely based out of 3 local hospitals in the Pacific Northwest region of the United States. As it is well known that safety, culture, and approach varies from state to state in the U.S, future studies should contain participants from other geographical areas of the country. Further research needs to investigate whether the effect of alarm modifications will bring similar benefits under other patient care settings such as intensive care, coronary care, emergency wards and medical –surgical units, etc. The entire experiment was executed in a simulator lab setting, which is very controlled and supported. As with any research work, the applicability of results from an experiment conducted in a controlled laboratory setting to a real life situation, which may be chaotic if it is a progressive care setting, needs to be examined further and may have to be repeated before being made into policies and procedures. A standard protocol or guidance for alarm modification will need to be developed for each unit in consultation with hospital administration and patient safety champions. Frequent modifications of alarms, without any baseline or guidance, may become a source of error and compromise patient safety.

In this study, direct cause-and-effect relationships could not be established between non-actionable alarms and contributors towards such alarms. To establish such relationships, other factors such as safety culture, hospital risk categories, technical characteristics of monitors used, and hospital protocols need to be included, and the study needs to be expanded into other departments. The issue of alarm fatigue is a system-wide challenge that needs to be approached holistically. This study included only one type of physiological monitoring device, and the results are applicable only to this type of monitor. Other types of monitors need to be studied in this setting before the results are generalized. The dependent variables in this study included response rate and error rate (any errors in the responded alarms). Although this is acceptable for a progressive care setting, response time (the time it took to address an alarm) and the severity of errors committed are critical in intensive/critical care units. Therefore, they should be studied in detail before the results are adopted.

Despite the significant reduction in alarm rate in this experiment, the primary issue behind alarm fatigue is training. Nurses and their assistants need to be appropriately trained to locate the user interface and adjust the setting so alarms can be reduced or eliminated. Therefore, a two-way comparison between alarm customization and training methods (classroom vs one-to-one) should be conducted. Without training, the benefit from alarm customization will be minimal. As the primary objective of this study is to reduce alarms that contribute to alarm fatigue, patient safety related outcome was not studied. The ultimate goal of any quality improvement project at a hospital is to enhance patient safety and reduce adverse events. Therefore, future studies should include and thoroughly study patient safety and patient satisfaction as dependent variables.

3.15 References

ACCE white paper on alarms. (2006). Retrieved from: http://thehtf.org/white%20paper.pdf. Last accessed: Nov 20, 2017.

Atzema, C., Schull, M. J., Borgundvaag, B., & Slaughter, G. R. et al. (2006) ALARMED: Adverse events in low-risk patients with chest pain receiving continuous electrocardiographic monitoring in the emergency department—a pilot study. *American Journal of Emergency Medicine,* 24, 62-67.

Barker, S. J. (2002), "Motion-resistant" pulse oximetry: a comparison of New and Old models. *Anesthesia and Analgesia*, 95(4), 967–972.

Bell, L (2010). Monitor alarm fatigue. *American Journal of Critical Care*, 19(1), 38-50.

Billinghurst, F., Morgan, B., & Arthur, H. M. (2003). Patient and Nurse-Related Implications of Remote Cardiac Telemetry. *Clinical Nursing Research*, 12(4), 356-370.

Bitan, Y., Meyer, J., & Shinar, D., (2004). Nurses' reactions to alarms in a neonatal intensive care unit. *Cognition, Technology and Work*, 6(4), 239–246.

Bliss, J. P., Gilson, R.D., &Deaton, J. E. (1995). Human probability matching behavior in response to alarms of varying reliability. *Ergonomics*, 38(11), 2300–2312.

Blum, J. M., & Tremper, K. K. (2010) Alarms in the intensive care unit: too much of a good thing is dangerous: is it time to add some intelligence to alarms? *Critical Care Medicine*, 38(2), 451–456.

Boev, C. (2012). The relationship between nurses' perception of work environment and patient satisfaction in adult critical care. *Journal of Nursing Scholarship*, 44(4), 368-375.

Bonafide, C. P., Lin. R., Zander, M., & Graham, C. S. et al. (2015). Association between exposure to nonactionable physiologic monitor alarms and response time in a children's hospital. *Journal of Hospital Medicine*, 10(6), 345-351.

Borowski, M., Gorges, M., Fried, R., & Such, O. et al. (2011). Medical Device Alarms. *Biomed Tech (Berl)*, 56(2), 73-83.

Burgess, L. P. (2009). Alarm limit settings for early warning systems to identify at-risk patients. *Journal of Advanced Nursing*, 65 (9), 1844 – 1852.

Chambrin, M. C., Ravaux, P., Calvelo-Aros, D., & Jaborska, A., et al. (1999). Multicentric Study of Monitoring Alarms in the Adult Intensive Care Unit (ICU): A descriptive Analysis. *Intensive Care Med*, 25(12), 1360-1366.

Christensen, M., Dodds, A., Sauer, J., & Watts, N. (2014). Alarm setting for the critically ill patient: a descriptive pilot survey of nurses' perceptions of current practice in an Australian Regional Critical Care Unit. *Intensive Critical Care Nurse*, 30(4), 204-210.

Clark, T. J., & Yoder-Wise, P. S. (2015). Enhancing trifocal leadership practices using simulation in a pediatric charge nurse orientation program. *Journal of Continuing Education in Nursing*, 46(7), 311-317.

Cvach, M. (2012). Monitor alarm fatigue: an integrative review. *Biomedical Instrumentation & Technology*, 46(4), 268-277.

Dandoy, C. E., Davies, S. M., Flesch, L., & Hayward, M. et al. (2015). A team based approach to reducing cardiac monitor alarms. *Pediatrics*, 134(6), 1686-1694.

Dandoy, C. E., Davies, S. M., Flesch, L., & Hayward, M. et al. (2014). A team-based approach to reducing cardiac monitor alarms. *Pediatrics*, 134(6), 1686-1694.

Drew, B. J., Harris, P., Zegre-Hemsey, J. K, & Mammone, T. et al. (2014). Insights into the problem of alarm fatigue with physiologic monitor devices: a comprehensive observational study of consecutive intensive care unit patients. *PLos ONE*, 9(10), 1-23.

ECRI Institute. Top 10 health technology hazards for 2012. Retrieved from: https://www.ecri.org/press/Pages/Top-10-Health-Technology-Hazards-2012.aspx. Last accessed: Nov 18, 2017.

ECRI Institute. Top 10 health technology hazards for 2013. Health Devices. 2014, 41, 342–365.Retrieved from: https://www.ecri.org/Forms/Pages/ECRI-Institute-2013-Top-10-Hazards.aspx. Last accessed: Nov 17, 2017.

Edwards, M., & Moczygemba, J. (2004). Reducing medical errors through better documentation. *Heath Care Management (Fred)*, 23(4), 329-33.

Edworthy, J. (2013). Alarms are still a problem!. *Anesthesia*, 68(8), 791-796.

Edworthy, J., & Hellier, E., Fewer but better auditory alarms will improve patient safety (2005). *BMJ Quality and Safety Health Care,* 14 (3), 212 – 215.

Falk, A. C., & Wallin, E. M. (2016). Quality of patient care in the critical care unit in relation to nurse patient ratio: A descriptive study. *Intensive Critical Care Nurse*, 35, 74 – 79.

Fast facts (2017). Fast facts on the U.S. hospitals. Web publication by AHA. Retrieved from https://www.aha.org/statistics/fast-facts-us-hospitals. Last accessed: Nov 20, 2017.

Funk, M., Clark, T., Bauld, T. J., & Ott, J. C. et al. (2014). Attitudes and practices related to clinical alarms. *American Journal of Critical Care*, 23(3), 9-18.

Funk, M., Parkosewich J. A., Johnson, C. R., & Stukshis, I., (1997). Effect of dedicated monitor watchers on patient outcomes. *American Journal of Critical Care,* 16, 318-323.

Gazarian, P. K. (2014). Nurses' response to frequency and types of electrocardiography alarms in a non-critical care setting: A descriptive study. *International Journal of Nursing Studies,* 51(2), 190-197.

Getty, D., Swets, J., Pickett, R., & Menditto, S. (1995). System operator response to warnings of danger: a laboratory investigation of the effects of the predictive value of a warning on human response time. *Journal of Experimental Psychology: Applied*, 1(1), 19–33.

Graham, K. C., & Cvach, M. (2010). Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms. *American Journal of Critical Care*, 19(1), 28- 34.

Hannibal, G. B. (2011), "Monitor alarms and alarm fatigue," *AACN Advancements in Critical Care*, 22, 418–420.

IEC 60601-1-8 (2006). Medical electrical equipment—Parts 1-8: General requirements for basic safety and essential performance—collateral standard: General requirements, test and guidance for alarm systems in medical electrical equipment and medical electrical systems. Geneva, Switzerland.

James, J. T. (2013). A new, evidence-based estimate of patient harms associated with hospital care. *Journal of Patient Safety*, 9(3), 122-128.

Konkani, A., Oakley, B., & Bauld, T. J. (2012). Reducing hospital noise: a review of medical device alarm management. *Biomedical Instrumentation & Technology*, 46(6), 478-487.

Korneiwicz, D. M., Clark, T., & David Y. (2008). A national online survey on the effectiveness of clinical alarms. *American Journal of Critical Care*, 17(1), 37–41.

Lawless, S.T. (1994). Crying wolf: false alarms in a pediatric intensive care unit. *Critical Care Medicine*, 22(6), 981-985.

Manzey, D., Gerard N., & Wiczorek, R. (2014). Decision-making and response strategies in interaction with alarms: the impact of alarm reliability, availability of alarm validity information and workload. *Ergonomics*, 57(12), 1833–1855.

Mitka, M. (2013). Joint Commission Warns of Alarm Fatigue: Multitude of alarms from monitoring devices problematic. *Journal of American Medical Association*, 309(22), 245-257.

Nix, M. (2015). Combating alarm fatigue. *American Journal of Nursing*, 115(2):16, 23-32.

Oliver, K., Innvar, S., Lorenc, T., & Woodman, J. et al. (2014). A systematic review of barriers to and facilitators of the use of evidence by policymakers. *BMC Health Services Research*, 14(2), 1-12.

Otero, A., Félix, P., Barro, S., & Palacios, F. (2009). Addressing the flaws of current critical alarms: a fuzzy constraint satisfaction approach. *Artificial Intelligence in Medicine*, 47(3), 219–238.

Pelter, M. M & Drew, B. (2015). Harm from alarm fatigue. Web M&M Case& Commentaries. Retrieved from: https://psnet.ahrq.gov/webmm/case/362/harm-from-alarm-fatigue. Last accessed: Nov 20, 2017.

Sanderson, M. P., Wee, A., & Lacherez, P. (2006). Learnability and discriminability of melodic medical equipment alarms. *Anesthesia*, 61(2), 142-147.

Sendelbach, S. & Funk, M. (2013). Alarm Fatigue: A patient safety concern. *Advanced critical care*, 24(4), 378-386.

Sendelbach, S., Wahl, S., Anthony, A., & Shotts, P. (2015). Stop the noise: A quality improvement project to decrease electrocardiographic nuisance alarms. *Critical Care Nurse*, 35(4), 15-22.

Shekelle P. G., Pronovost P. J., & Wachter R.M. (2011). Advancing the science of patient safety. *Annals of Internal Medicine*, 54 (10), 693 – 696.

Siebig, S., Kuhls, S., Imhoff, M., & Gather, U. et al. (2010). Intensive care unit alarms – how many do we need? *Critical Care Medicine*, 38, 451–456.

Sowan, A. K., Gomez, T. M., Tarriela, A. F., & Reed, C. C. et al. (2016). Changes in default alarm settings and standard in-service are insufficient to improve alarm

fatigue in an intensive care unit: A pilot project. *JMIR Human Factors*, 3(1), 28-35.

Sowan, A. K., Tarriela, A. F., Gomez, T. M., & Reed, C. C. et al. (2015). Nurses' perceptions and practices toward clinical alarms in a transplant cardiac intensive care unit: Exploring key issues leading to alarm fatigue. *JMIR Human Factors*, 2(1), 45-52.

Spence, H. K. L., & Leiter, M. P. (2006). The impact of nursing work environments on patient safety outcomes: the mediating role of burnout/engagement. *Journal of Nursing Administration*, 36(5), 259 – 267.

Taenzer, A. H., Pyke, J. B., & McGrath, S. P. (2011). A review of current and emerging approaches to address failure-to-rescue. *Anesthesiology*, 115, 421–431.

Talley, L.B., Hooper, J., Jacobs, B., & Guzzetta, C. et al. Spring 2011. Cardiopulmonary monitors and clinically significant events in critically ill children. *Horizons*, 38-45.

The Joint Commission (2013). The Joint Commission sentinel event alert: medical device alarm safety in hospitals. Retrieved from http://www.jointcommission.org/assets/1/18/SEA_50_alarms_4_5_13_FINAL1.PDF. Last accessed: Oct 1, 2016.

The Joint Commission (2015). National patient safety goals effective January 1, 2015. Retrieved from http://www.jointcommission.org/assets/1/6/2015_NPSG_HAP.pdf. Last accessed: Oct 1, 2016.

Tsien, C. L., & Fackler, J. C. (1997) Poor prognosis for existing monitors in the intensive care unit. *Critical Care Medicine*, 25, 614-619.

Varpio, L., Kuziemsky, C., MacDonald, C., & King, W. J. (2012). The helpful or hindering effects of in-hospital patient monitor alarms on nurses: A qualitative analysis. *Computers, Informatics, Nursing*, 30(4), 210-217.

Welch, J. (2011). An Evidence-based Approach to Reduce Nuisance Alarms and Alarm Fatigue. *Biomedical Instrumentation & Technology*, 45(1), 46-52.

Whalen, D. A., Covelle, P. M., Piepenbrink, J. C, & Villanoza, K. L. et al. (2014). Novel approach to cardiac alarm management on telemetry units. *Journal of Cardiovascular Nursing*, 29(5), 13-22.

CHAPTER IV

THE RELATIONSHIP BETWEEN WORK LOAD INDEX AND ALARM RESPONSE RATE AND ERROR RATE

4.1 Abstract

Objective: The primary objective of this study was to assess the perceived workload index of nurses/assistants providing patient care under different medical device alarm settings. Furthermore, this study also attempted to establish the relationship between the workload index and performance that was measured by the number of alarms addressed and errors committed while caring for patients.

Methods: Thirty participants, 23 females and 7 males, responded to alarms that occurred on a physiological monitor under two conditions (default and modified) for a given clinical condition in a clinical simulator. Cumulative alarm response rate, which is the number of alarms responded to out of a set total presented, was recorded for each condition. Along with alarm response rate, number of errors committed (i.e. error rate) were also recorded for each alarm condition. Upon completion of all assigned tasks, study participants completed a NASA-TLX questionnaire (on an iPad ® application) for each condition. NASA-TLX was used to measure the subjective dimensions of mental demands, physical demands, temporal demands, effortfulness, personal satisfaction with job performance, and frustration level during work for each setting. The study participants rated the demand experienced on a 20-point visual analogue scale with

anchors of 'low' and 'high' for each subscale. The overall workload measurement was then obtained by summing the weighted scores on the subscale. Between-subjects ANOVA was performed for the various dependent variables at different alarm settings. In addition, Pearson correlational analysis was performed between the workload score and alarm response rates and error rates to establish relationships.

Results: Study participants experienced lower workload when the medical device alarm threshold limits were modified according to patients' clinical conditions. Significant correlations were found between the workload index and measured dependent variables – alarm response rate and error rate. Higher alarm workload corresponded to a higher number of committed errors in alarm management and a lower number of addressed alarms. Adversely, a lower alarm workload, presented by customizing the alarm settings, resulted in a lower number of committed errors and a higher number of alarms addressed. The perceived workload index was comparatively lower in an environment with alarm settings modified for individual patient care, than in a patient care environment where the medical equipment operated under default settings.

4.2 Introduction

Nurses have complained about high levels of workload in their field in terms of the amount of patient care activity they must perform (Myny et al., 2011). In turn, higher nursing workload is considered a contributor to sentinel events and poor patient care quality (Bogaert et al., 2013). Characteristics of the work environment and complexity of the work system are key factors in the nurses' increased workload. Overwork, fatigue, incorrect physician order, handoff communication with patients and other units, and problems with medical devices present in the environment pose serious threats to overall

patient outcome and safety (Feder and Funk, 2013). Of these identified factors, medical devices are expected to aid the nurses rather than cause and contribute to fatigue. High nurse-to-patient ratios impact staff performance levels and obstruct staff's ability to respond to devices, each of which carries its own parameters and operational characteristics.

Generally, excessive workload has been associated with stress, patient safety, outcomes, and performance decrements in intensive care work environments. Although fatigue and workload are conceptually different, they are closely related. Soh and Crumpton, in their landmark study, described fatigue as a multi-causal, multi-dimensional, non-specific and subjective phenomenon resulting from prolonged activity and psychological, socioeconomic and environmental factors that affect both the mind and the body (Soh and Crumpton, 1996). Fatigue can come directly from job demands such as work schedule, workload and extended patient care hours. It is ironic that the devices and alarms created to support nurses and reduce workload level contribute to fatigue and increased workload. Nurses are a very important resource who directly affect the healthcare system; therefore, providing optimal workload level is imperative (Zboril-Benson, 2002).

Workload level and sources of stressors have been implicated as sources of error not only in healthcare settings, but also in multiple other settings. Research shows that workload is one of the most important job stressors among critical care unit and intensive care unit nurses (Wolf et al, 2006). Generally, researchers in aviation and nuclear power plants study the relationship between workload and human performance. However, unlike other industries, nursing workload is more than just the number of tasks required

or output of a nurse in the course of an 8-hr or 10-hr shift. Nursing workload, especially while caring for patients, is complex and somewhat nonlinear by nature (Zboril-Benson, 2002). Nursing workload is often thought of as more mentally than physically demanding work. Although some studies have reported physical demands associated with nursing work tasks, a significant number of tasks are mentally taxing (Szczurak et al, 2007).

Routine nursing tasks such as communicating with patients, answering care related questions, administering medications, reviewing medication lists, answering pharmacy or other providers' calls and managing medical equipment require expertise in cognitive skills like making judgements, accommodating memory demands, and managing mental workload (Hyde et al, 2009). Thus, mental and physical fatigue are both likely present among nurses, making it necessary to understand the consequence of elevated workload while caring for patients.

A significant number of medical devices present in typical patient care settings are physiological monitors (Grossman et al., 2011). It is a standard nursing practice to rely on these devices to continuously watch patients when caring for other patients. The medical devices present in the work environment alert nurses when deviations occur from a preset limit. When the number of alarms to which nurses must respond far exceed nurses' capacity to respond, they encounter fatigue and fail to respond efficiently to every alarm (Borowski et al., 2011; Way et al., 2014). Currently, there is no consensus among researchers on the definition of alarm-associated fatigue (Deb & Claudio, 2017). However, it has been shown that alarm associated fatigue due to 'cognitive-information overload' – receiving so much information and so many demands that the human brain cannot process and handle them all – desensitizes nurses and leads them to feeling

burdened (Konkani et al, 2012). Alarm fatigue attributed to desensitization may result in slower nursing response time and may cause nurses to ignore or turn off alarms altogether, defeating the purpose of alarm design (Welch, 2011; Burgess et al., 2009). It is common for nurses to disable alarms when they cannot respond (Kanwar et al., 2008; ECRI 2014).

Many devices found in intensive care units have alarms, including patient call systems, infusion pumps, ventilators, emergency resuscitation devices, pulse oximeters, apnea monitors and other life support equipment (Drew and Funk, 2006; Billinghurst et al., 2003). They are designed to care for patients who are seriously injured, have a critical or life-threatening illness, or have undergone a major surgical procedure, thereby requiring 24-hour monitoring (Johnson et al., 2017; Sachdev et al., 2010; Harris et al., 2011). Some of these devices may alarm simultaneously, resulting in alarm related burden and cognitive information overload (Harris et al., 2011; Chopra and McMohan, 2014).

To date, little research has been conducted in the area of workload and its correlation to alarm hazards and nurse response time. Although a number of researchers have reported that nurses' fatigue contributes to alarm mismanagement, no studies have been performed to quantify the fatigue during alarm management and its effect on patient care quality and outcome. The National Aeronautics and Space Administration Task Load Index (NASA-TLX) provides a subjective measure of mental demand, physical demand, and temporal demand, the subjects' own performance, effort and frustration (Hart and Staveland, 1988). Overall workload measurement is then obtained by summing the six subscales. While some researchers have assessed the mental workload in a clinical

setting, they have not addressed the specific impact of increased workload on alarm management, response rate and error rate (i.e. number of alarms addressed incorrectly while providing patient care) (Hoonakker et al., 2011; Yamase, 2003; Carayon and Gurses, 2005). The aim of this study is to understand how the subjective and objective levels of mental workload influence nurses' performance as a function of situation complexity and alarm management experience. In other words, this study intends to verify whether any changes in situational complexity increase the subjective and physiological levels of mental workload and lead to any performance issues while providing nursing care. This study will examine whether any relationship exists between nurses' mental workload and alarm response and error rate.

4.3 Nursing workload and its impact

Hospital operators have increased the number of beds due to a spike in demand not only in intensive care units, but also across the spectrum -- in recovery units, pre - and post-operative units, step-down care units and emergency wards in the United States. This is commensurate with changes in demographics, progress in diagnostics, and therapeutic methods that lead to the prolongation of a patient's life (Kaminski et al., 2015). Nursing staff and their assistants provide the majority of care at these bedsides. Shortages of skills and resources contribute to increased mental workload for the existing nurses (Young et al., 2008). Factors like staff and skill shortage will increase fatigue and mental workload within this occupational group among healthcare professionals (Dye & Wells, 2017).

Excessive workload (both physical and mental) is a major contributor to work related stress in nurses. Jobs with a high level of workload and occupations with

constantly changing work schedules diminish nurses' performance and result in irritability, decreased productivity, impaired decision-making capacity and reduced ability to learn new concepts. Errors nurses commit during diagnosis of a problem or making of clinical decisions are more often due to cognitive errors than a lack of basic knowledge (Brown et al., 1997). Due to their training and licensing requirements, healthcare professionals such as nurses and doctors typically possess an adequate knowledge base. Very often, missed clinical steps or wrong decisions by healthcare professionals during treatments are due to mental fatigue (Hravnak et al., 2011). Mental fatigue is typically characterized by exhaustion and reduced interest in task execution.

In recent years, medical errors have received a great deal of attention in the United States. According to an ECRI Institute report on incident and recall analysis, significant medical errors occur during medication administration (ECRI, 2014). Administering infusions intravenously is the most common practice in medication administration (Brown et al., 1997). Errors made during programming, troubleshooting, addressing alarms and preventative maintenance of the infusion pump can have dramatic consequences (Rosman et al., 2013). According to researchers, one of the primary causes of errors while operating infusion pumps is increased mental workload induced by excessive alarms presented by the infusion pump (Varpio et al., 2012). Device manufacturers include alarms, alerts and warnings to notify nurses when there is a change in machine or patient status for early detection of abnormalities. Several types of devices -- infusion pumps, physiological monitors, and therapy delivery devices -- are used in typical patient care settings, and multiple alarms from these devices can cause information overload, leading to clinical errors and poor overall patient outcomes. In

alarm management, nurses perform many activities that require excessive cognitive processing, which may contribute to sensory overload, and as a result, their alertness may decrease and human errors may occur (Taenzer et al., 2010). In particular, mental overload may decrease functioning of working memory. Therefore, it is important to assess the mental workload in attending nurses while they are operating these medical devices during patient care.

Many ergonomists and researchers have applied subjective and physiological measures to evaluate mental workload quantitatively in healthcare and other fields. Given the fact that nursing is a complex field, cognitive workload cannot be described using one dimension or characteristic. According to Neill (2011), an individual's processing capacity is affected by work and personal related factors including environmental and organizational factors as well as perception. Due to rampant use of technology, work requirements for nurses have shifted from the physical to the mental realm. A few other researchers have also concluded along the same lines. Specifically, Veltman (2002) proposed that mental workload techniques could be grouped into three broad measures: psychophysical, performance and subjective. Each of these measures has specific applications and limitations in determining the mental workload associated with the work demands. In subjective mental workload, the worker knows the amount of work needed to meet a particular demand. Subjective workload scales have been a familiar part of the human factors and ergonomics toolkit since the 1980s (Tsang & Vidulich, 2006). Although workload metrics can also be obtained from both performance and physiological measures, subjective scales have the advantage of accessibility, ease of use, and direct applicability to situations where the operator's experience is of paramount

concern. Few studies have evaluated the impact of nurses' workloads on the nature of care provided and patient outcomes (Tubbs-Cooley et al., 2014).

Multiple factors such as work environment, time pressure to complete the task, and the individual's prior experiences influence an individual's perception of workload. Ead (2015) described a framework of nursing workload that incorporates both exogenous and endogenous variables. Exogenous variables are external factors that include the complexity of the patient, staffing resources, and deviations from daily routine. Endogenous variables are internal to the organization and typically include the nurse's own coping ability, knowledge, and demand predicting ability, energy level, and organizational skills. Development of conceptual frameworks specific to alarm management and research regarding alarm related workload is lacking. This research will assess workload when addressing alarms and the cascading effect the alarms have on nurses' and their assistants' other primary tasks (e.g. patient assessments, medication administration). Because today's healthcare environment is a multi-tasking system, time and effort spent responding to alarms detracts from nurses' primary tasks. As primary task workload increases, alarm task performance typically worsens, particularly when alarm reliability is low. Alarm response rate in this situation may be low because the operator must choose an action based on relative urgency of the primary and secondary tasks (Gomez et al., 2015). Differentiated alarm tones could help nurses in assessing the need for prioritizing a visit to the source of the alarm (e.g. physiological monitor or infusion pump) versus completing a primary task followed by attention to the alarm.

4.4 Alarm fatigue, nursing and patient safety

Medical device alarms and alerts, specifically designed by medical device manufacturers, are intended to alert clinicians to any deviation of physiological signals from the normal value. Although the intention sounds appropriate and ensures that doctors and nurses will always be informed of physiologic changes in order to respond to important deterioration events quickly, we know that these devices generate very frequent alarms, and that a significant proportion are false (Liu & Pecht, 2011; Funk et al., 2013). Clinical alarm system safety has received immense attention from clinicians, hospital administrators, and watchdog agencies especially after the U.S. Food and Drug Administration's (FDA) report indicated 566 alarm related patient deaths (The Joint Commission 2013; www.jointcommission.org/assets/1/18/SEA_50_alarms_4_5_13, accessed 10-10-2017). The task of separating the true, actionable alarms from the false or non-actionable alarms falls to the clinicians responsible for responding to alarms, who in most settings are nurses (Dressler et al., 2014).

Medical device alarms system safety is complex. Alarm fatigue among health care workers, especially nurses, poses a risk to patient safety (Gazarian, 2014; Buist et al, 2004). Upon deciding and initiating appropriate medical treatment, doctors handoff patients from their care to nurses and their assistants while recovering. Patients need to be continuously monitored during this recovery phase for any changes in status (Burgess et al, 2009). When caring for multiple patients, nurses are exposed to numerous alarms per patient per shift and over time become fatigued due to an overwhelming amount of alarms (Gross et al., 2011). One solution frequently suggested to reduce fatigue is to adjust alarm parameters to suit patient conditions or a standard hospital protocol rather

than using textbook normal values. According to Turmell et al. (2017, pp 48-49), "alarm fatigue is composed of 2 components - the first is alarm desensitization that stems from excessive alarms causing the nursing staff to tune out the alarm or silence it and the second component to alarm fatigue is alarm apathy." Physiological monitors and medical devices generate frequent alarms and most are not relevant for making clinical decisions, providing patient care or ensuring patients' safety. By one estimate, 70% of alarms occurring in adult intensive care units are not adding any value to the nurses' work process when monitoring patients (Pergher & Silva, 2013).

Many medical device manufacturers set the alarm threshold values to text normal values and increase the sensitivity of equipment as much possible, which results in an excessive number of non-actionable and false positive alarms (Liu & Pecht, 2011). Over the course of few shifts or few days, nurses become desensitized to these alarms and start devaluing them. This insensitivity towards medical device alarms can result in harmful patient safety related consequences. While alarm fatigue has been recognized as a threat to patient safety, and studies have been conducted to assess its impact on patient care, the phenomenon of alarm fatigue has yet to be fully quantified in a subjective or objective way (Deb & Claudio, 2017). The literature search revealed no studies on nurses' mental workload assessment while managing medical device alarms. This research effort will be the first to assess nurses' mental workload under different alarm management settings (default and modified). Data collected from this study could be used to develop procedures, framework and policies for managing medical device alarms and developing solutions to reduce alarm fatigue.

4.5 Importance of evidence based practice

As the demand for high quality healthcare rises, nurses are expected to deliver patient care at levels higher than they are trained to operate as part of degrees. Weng et al. (2015) define evidence-based practice as "a process of collection, interpretation, appraisal, and integration of valid, clinically significant, and evidence-based implementation". Today, across the globe, most major healthcare organizations and watchdog agencies push for better patient outcomes and strive to improve quality and consistency of care through integration of evidence-based practice (Munten et al., 2010; Saunders & Vehviläinen-Julkunen, 2016). The implementation of evidence-based practice (EBP) in healthcare is presented as the panacea to the all issues and challenges faced by nursing.

The Institute of Medicine (IOM) focuses on improving health care quality in order to increase positive outcomes through consistent use of research-based knowledge (Physician-Patient Alliance for Health and Safety, 2013). All healthcare providers, hospitals, government regulatory entities and non-profit watchdog agencies can be held financially and legally accountable for patient related adverse events as policies are established based on evidence-based nursing (Stevens, 2013). According to Brower (2017, pp 16), "the spotlight on EBP is certain to increase as evidence continues to be produced, healthcare legislation changes, and consumer demands for quality healthcare and accountability increase." This will result in system-wide change that would affect the general population in a positive way. The ultimate goal of nursing practice is to improve health outcomes and make patients' lives better. A key challenge in this time of lean operations in the name of efficiency and nursing shortage is keeping up with advances.

Therefore, it is often easier to rely on traditional nursing practices, despite the availability of evidence from a body of knowledge. Sackett et al. (2000) define evidence-based practices as the integration of best research evidence with clinical expertise and patient values. In broad terms, best research evidence refers to scientifically sound, clinically relevant research, uncompromised safe and clinical effectiveness, and remarkable patient outcomes. Research findings, knowledge from basic science, clinical knowledge, and expert opinion are all considered "evidence"; however, practices based on research findings are more likely to result in the desired patient outcomes across various settings and geographic locations. Evidence-based practice also provides opportunities for nursing care to be more individualized, more effective, streamlined, and dynamic, and to maximize effects of clinical judgment (Saunders & Vehviläinen-Julkunen, 2016). When evidence is used to define best practices rather than to support existing practices, nursing care keeps pace with the latest technological advances and takes advantage of new knowledge developments. The process of implementing evidence-based practices begins by recognizing a clinical concern, generally from widely recognized bodies such as The Joint Commission, FDA and ECRI, that can be solved through application of evidence. According to Melynk and Fineout-Overholt (2015), completing pilot tests, collecting scientifically sound data, and comparing and contrasting of pilots' test outcomes against the current method are key steps in implementing evidence-based practice. This research effort essentially follows the guidelines proposed by these researchers -- conducted experiments using nurse participants, collected data, compared to the existing practice, and critically evaluated the results to assess benefits.

4.6 Hypotheses

The experimental hypotheses are defined as follows:

1. Modified alarm setting will have lower workload index than default alarm setting.
2. Higher alarm workload corresponds to a higher number of committed errors during alarm management.
3. Higher alarm workload environment corresponds to lower response rate. That is, the higher the alarm workload, the fewer the number of alarms that will be addressed.
4. Customizing alarm threshold will result in lower mental workload and lower error rate.
5. Lower mental workload in nurses corresponds to better patient care and overall satisfaction.

4.7 Methods

In order to establish the correlation between alarm management workload and number of committed errors and response rate, the experimental methods and participants are defined as follows. The independent and dependent variables in this section reflect the hypotheses listed above.

4.7.1 Experimental design

A between subjects ANOVA was used to test for differences in mental workload, error rate and response rate between modified and default alarm threshold settings. Although this study was conducted immediately after the experiment # 2 (previous

study), it was independent in nature. As such, it used the same participants and data collected – error rate, response rate, overall satisfaction and patient care experience and established their relationship to subject workload assessment ratings. Mental workload was measured using subjective assessment ratings utilizing an electronic version (iPad® application) of the NASA-TLX index. As stated in the previous study, error rates were measured by the number of alarms addressed incorrectly, and response rates were calculated by the number of alarms addressed against the total number of alarms presented. NASA-TLX ratings were collected at the end of each segment (modified and default settings) using an iPad application. The study subjects were randomly assigned to one of the two alarm-setting groups – normal setting group or modified alarms group.

4.8 Variables

The primary independent variable in this study was level of alarm setting: default and modified. The default alarm and alerts setting, as the name indicates, is set by the device manufacturer at the time of release to the market. The hospital's biomedical staff and administrators put the device on the treatment floor without any alterations to the limit. The modified setting is set by (usually Biomedical department staff) incorporating rules and algorithms for alarms and alerts with a goal of reducing non-actionable and nuisance alarms. The number of alarms for each setting is shown in Table 4.1.

Dependent variables for this study included subjective workload (assessed via NASA TLX method), error rate, response rate, overall satisfaction and patient care experience. The procedures used to collect data for this study were identical to those in the previous study. A brief summary is provided in subsequent sections. Upon finishing the experiment for the previous study, each participant filled out an electronic version of

NASA-TLX. The NASA-TLX provides a multidimensional rating procedure that allows for collecting subjective workload scores based on a weighted average of ratings of six subscales or raw scores ("Raw TLX Scores"). The six subscales include: Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Own Performance (OP), Effort (EF), and Frustration Level (FL). According to Hart (2006), Raw TLX is simpler to use and gives similar results for the total mental workload score as weighing method. Therefore, Raw TLX method was used to assess subjective workload. The mobile application version of the tool was downloaded onto a mobile computing media (iPad) from the NASA website (https://humansystems.arc.nasa.gov/groups/tlx/tlxapp.php; last accessed: Aug, 2017) and used for this study. Error rate is defined as the percentage of incorrectly addressed alarms in a setting. Response rate is defined as the percentage of addressed alarms, correctly as well as incorrectly, during the experiment(s).

[Intentionally left blank]

Table 4.1 Number of alarms

Items	Default Setting	Modified Setting
Total number of 'actionable' alarms	2	2
Total number of 'non-actionable' alarms	13	6
Total number of alarms to be addressed by each participant	15	8

4.9 Participants

Participants consisted of the same 30 nurses and their assistants from the previous study. The sample was composed of 23 males and 7 females, ranging from 24 to 60 years of age (M = 40.67, SD = 9.85), based on a random sample of the nurses from various local area hospitals. Flyers and word-of-mouth were used to recruit participants. Participants were randomly assigned to one of two alarm-setting groups – normal setting group and modified setting group, and each group included 15 participants. Inclusion criteria for the previous study required some medical alarm exposure. There were no exclusion criteria for this study.

Previous studies on mental workload in healthcare settings were examined to determine the appropriate sample size. A study by Holden et al. (2010) utilized 79 pharmacy professionals in assessing the mental workload demands during medication dispensing and administration. Sample sizes for previous workload impact assessment studies in intensive care unit (ICU) and healthcare settings ranged from N = 16 (Malacrida, et al., 1991) to N = 81 (Mohammadi, et al., 2016). Related studies examining healthcare professionals' performances in high workload environments where the researchers used a sample size of N = 12 and N = 31 (Abelson, et al., 2016) were also

considered. Therefore, it was determined that a sample size of 30 would suffice for this experiment.

4.10 Experimental procedure

Participants were seated at a workstation desk in a conference room setting. The data collection session began with an overview of the process and experiment, and participants were reminded of the previously completely informed consent form (Appendix A). The participants were given time to review the previously completed informed consent form and ask for any clarifications.

Participants were briefed verbally as to the purposes of the experiment and given printed instructions for completing the NASA-TLX electronic survey. The NASA-TLX was used to characterize the workload that the participants were subjected to by each of the two alarm threshold settings – modified and default – while providing patient care. Immediately upon finishing the tasks in the previous study, which included alarm management on an instrumented mannequin and associated other tasks in a progressive care setting, study participants completed the NASA-TLX electronic survey for this experiment. Before scoring, each participant was trained on the connotation of the six-subscales of the NASA-TLX method, and also familiarized with how to use the computer version for scoring. The study participants were asked to mark on the twenty-step bipolar (low to high/good to poor) subscale by touching the tick mark location based on their perception of the contribution of that particular subscale to the workload of the alarms. The rating scale definitions for these six dimensions are shown in Appendix A. All the participants rated their experience for alarm management on 20-step scales for each of the six dimensions. The vertical tick marks on each sub-scale divide the scale from 0 to

100 in increments of 5. If a participant marked in between tick marks, it was rounded to the nearest 5 to the right of the marking. Similar to the previous study, the four administrative tasks were not included for data collection and analysis. An overall mental workload score for NASA-TLX was obtained on a scale of 0 through 100 for each participant by calculating an average of those six ratings for six dimensions.

4.11 Statistical analysis

Descriptive statistics for all workload measures were computed, and normality tests were performed as appropriate. Correlations between the workload and participants' alarm response were determined by calculating the Pearson product-moment correlation coefficient. Correlations between the workload and other dependent variables such as committed error rate, care provide experience and overall satisfaction were determined using Spearman's correlation method. Correlation coefficient ranges for positive relationship were defined as follows: $r < 0.3$ as a weak correlation, $0.5 > r > 0.3$ as a moderate correlation, and $r > 0.5$ as a strong correlation. The same ranges but with negative magnitude were considered as inverse correlations.

A series of comparison tests of $\chi 2$ were performed to examine if subscales' scores differed as a function of demographic characteristics (i.e., age, gender, years of experience as a nurse, and alarm management experience). No differences were noted across analyses ($p > .05$). All results were considered significant at an alpha level of 0.05. The SPSS statistical software package version 9.2 for Windows was used for all statistical analysis.

4.12 Results

The demographic data variables of age, alarm management experience, gender, education and alarm management training were collected at the beginning of the testing session and were classified and described in Study 2. Descriptive statistics for the dependent variables (i.e. alarm response rate, committed error rate, care provider experience and overall satisfaction) are shown based on alarm settings in Table 3.3, and descriptive statistics for workload measures are shown in Table 4.2. Overall, the variables were found to be significantly different between default and modified alarm settings.

Table 4.2 Descriptive statistics for six subscales and overall score

	Average (SD)		
Subscale	Default Setting	Modified Setting	Total Participants
Mental Demand (MD)	65.3 (8.5)	45.7 (8.2)	30
Physical Demand (PD)	32.3 (5.3)	30.7 (6.5)	30
Temporal Demand (TD)	75.0 (7.3)	60.7 (6.5)	30
Own Performance (OP)	53.3 (7.2)	66.0 (8.1)	30
Effort (EF)	50.3 (6.7)	51.7 (7.7)	30
Frustration Level (FL)	69.3 (7.5)	59.7 (8.5)	30
Overall	57.6 (2.6)	52.4 (2.3)	30

4.12.1 Workload index

An independent-samples t-test was run to determine if there were differences in participants perceived workload between modified and default settings. There were no outliers in the data, as assessed by inspection of a boxplot. Workload index scores for each of six subscales were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there was homogeneity of variances, as assessed by Levene's test for equality of

variances (p = .18). The workload index was higher for the default alarm setting (57.60 ± 2.59) than the modified alarm setting (52.39 ± 2.29), a statistically significant difference of 5.21 (95% Confidence Interval, 3.38 to 7.04), t (28) = 5.838, p < .05. Participants' individual ratings for each subscale along with computed overall workload index is shown in Figure 4.1.

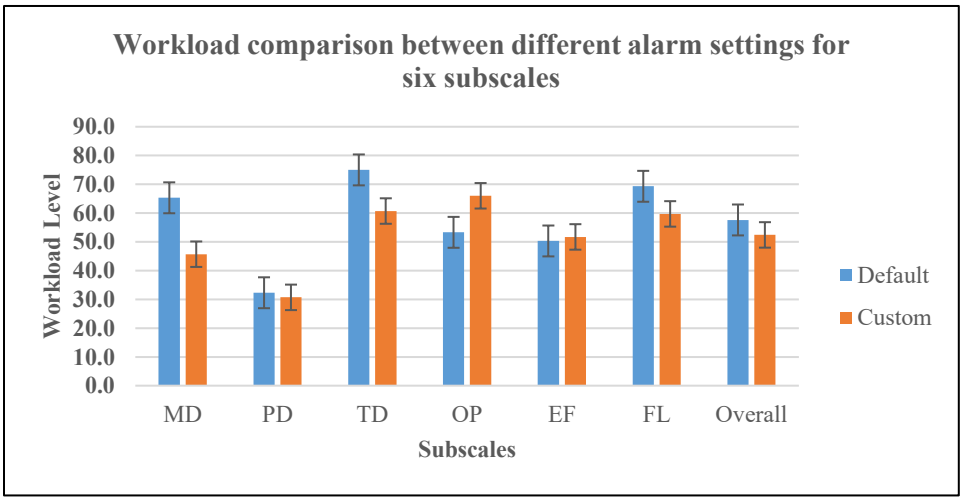

Figure 4.1 Subscale comparison chart for different alarm settings

4.12.2 Relationship between alarm workload and alarm response rate

A Pearson's product-moment correlation was run to assess the relationship between workload and the number of alarms addressed (alarm response rate) while providing patient care. Analyses showed the relationship to be linear with both variables normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there were no outliers. There was a strong negative correlation between alarm response rate and perceived workload, r (28) = -.54, $p < .002$, with workload explaining 29% of the variation in alarm response rate. The negative correlation indicates that an increase in alarm workload is associated with a reduction in number of addressed alarms. In other

words, modification of alarms, according to patient conditions, in patient supporting medical devices would help reduce workload for care providers and improve alarm response rate.

4.12.3 Relationship between alarm workload and alarm error rate

A Spearman's rank-order correlation was run to assess the relationship between alarm error rate and perceived workload while providing patient care. Analysis showed the relationship to be monotonic, as assessed by visual inspection of a scatterplot. There was a strong positive correlation between the number of errors committed (alarm error rate) and the perceived workload, $r_s(28) = .60$, $p < .05$. The number of errors committed by nurses/assistants dropped simultaneously with corresponding workload, which shows that they associated with each other in a healthcare environment.

4.12.4 Relationship between alarm workload and care provider experience

A Spearman's rank-order correlation was run to assess the relationship between perceived workload and care provider experience while providing patient care in a progressive care setting. Analysis showed the relationship to be monotonic, as assessed by visual inspection of a scatterplot. There was a moderate negative correlation between the experience reported during patient care and the perceived workload, $r_s(28) = -.49$, $p < .05$. The care provider experience, when or after caring for patients, was found to be inversely proportional to the alarm related workload. It is important to note that the study participants were managing alarms in addition to several patient care tasks to mimic real world situations. Therefore, any reduction in workload positively impacted care provider experience and well-being at the job.

4.12.5 Relationship between alarm workload and overall satisfaction

A Spearman's rank-order correlation was run to assess the relationship between perceived workload and care provider experience while providing patient care in a progressive care setting. Analysis showed the relationship to be monotonic, as assessed by visual inspection of a scatterplot. There was a strong negative correlation between the overall reported satisfaction and perceived workload, r_s (28) = -.69, p < .05. The negative correlation indicates that the workload increase is associated with the overall satisfaction, which dropped significantly. Therefore, hospital administrators and risk managers should seriously consider customizing alarms in patient-supporting medical products, as it is a key contributing factor in care provider's satisfaction.

4.13 Narrative data

A total of 5 nurses provided comments for open ended questions. Four responders in default setting provided narrative comments about alarm management and issues in timely completion of patient care tasks. All four comments were negative, reflecting an excessive number of alarms and an excessive amount of patient care tasks. The one comment provided by a participant in modified setting appeared to be positive. The comments are listed verbatim in Table 4.3

Table 4.3 Narrative data for open question (Q3 in A2)

Alarm setting	Comments
Default	"Stressful to manage alarms"
	"Too many tasks; crunched for time"
	"Somewhat intense alarms for a progressive care unit"
	"Rough unit it seems!. Too many things to do"
Modified	"Easy to work in here!"

4.14 General discussion

Delayed or no response to impending patient safety related calls, poor care provider experience, low job satisfaction, and adverse events are all unwanted outcomes of alarm fatigue. Nurses often cite increase in alarm related workload as a reason for alarm fatigue; it has shown to be a major contributor for aforementioned unwanted outcomes. Increased workload affects both the care provider and the patient. No studies to date have been conducted to measure the workload while caring for patients and managing alarms simultaneously, and relate that measurement to primary measures of alarm fatigue – response rate, overall satisfaction and care provider experience. The intent of this study was to provide alarm fatigue researchers some insights into the relationship between workload and key measures of alarm fatigue. This study measured the perceived workload under two different alarm settings and associated it to various alarm fatigue measures quantitatively.

The alteration of alarm limits based on patients' conditions by customizing experimental settings resulted in lower NASA-TLX scores compared to default manufacturer settings. In other words, allowing the physiological monitoring device to operate under a default setting that is based on textbook normal values resulted in more alarms, which in turn resulted in higher mental workload during management of these alarms. Higher NASA-TLX scores indicate that alarm management is a complex task and has the potential to induce fatigue. Higher mental workload impacts nurses' attentiveness, increases the risk of slow responses, and can result in poor task accuracy. In a study conducted by Cvach et al. (2013), the number of alarm signals reached several hundred per day for some patients, creating a high alarm burden for nurses. Nurses, due to high

alarm burden, will get desensitized and may miss, ignore or disable alarm signals, which might result in an adverse event.

The NASA-TLX provides an overall index of mental workload as well as the relative contributions of the six subscales: mental, physical and temporal demands, and effort, frustration and perceived performance. The subscales scores show that TD, MD and FL are large contributors to alarm workload. This is not surprising as responding to alarms is secondary to primary care provider tasks such as medication administration, patient assessments and note updates. In such dual-task systems, time spent responding to alarms distracts from the primary tasks, and nurses feel pressed for time and frustrated. The higher MD score is due to the process involved in analyzing and isolating the source of the alarm, which often requires higher cognitive amplitude.

The study participants' self-reported performance was higher in modified setting than in default setting. Higher alarm response rate in modified setting supports this score. Better alarm response rate is also manifested across two other subscales – as lower frustration and overall workload index as shown in Figure 4.1. Not surprisingly, the subscale scores for PD and EF for modified and default settings were statistically similar and lower than other subscales in their respective groups. Though only 13.3% of participants provided narrative data, making it difficult to generalize for the entire group, the common theme for default setting was that the number of alarms and tasks was excessive. The sole comment from a participant in modified setting was generically positive and did not provide any explicit information about alarm management or workload.

The most important finding from this study is that the number of alarms addressed was inversely proportional to the workload encountered during patient care. The study participants were able to address almost all of the presented alarms when the alarm settings were modified according to patient conditions. This finding is consistent with other findings in similar alarm setting modification studies. An initiative led by researchers Cvach et al. (2013) at Johns Hopkins Hospital demonstrated that a 43% reduction in alarms was possible through alarm setting customization. The study participants in this quality initiative project expressed positive views about alarm customization. Dandoy et al. (2014) reduced the total number of alarms from 180 per patient per day to 40 through a unit-level standardization project, which included a daily individualization of alarm parameters. In a study conducted by Srinivasa et al. (2017), the researchers permanently turned off three types of ventricular contraction alarms, creating a 54% decrease in the total rate of alarms per bed per day and a significant noise reduction in the units.

Another unique finding from this study was that the alarm workload was directly proportional to the number of errors committed. The drop in number of errors committed is associated with the number of alarms that needed to be addressed during patient care. This direct relationship suggests that the removal of certain non-essential alarms enabled the nurses to address the remaining important alarms accurately without any or minimal errors. The nurses had more time and spent that time in addressing the presenting alarms appropriately. The overwhelming number of alarms in default setting put time pressure on nurses and thus, they attempted to address more alarms within the limited time and made errors along the way. This can also be seen in a different way – if the number of

opportunities (alarms) to make an error is limited, the number of committed errors will likely reduce.

Care provider experience and overall satisfaction were found to be inversely correlated to alarm related workload. As the alarm related workload increases, which is typical when the alarms are set at the manufacturer's default setting, the quality of experience for care providers caring for patients decreases. When the number of alarms to be assessed and addressed are low(er), nurses and their assistants have more time to focus on patient care tasks and support other critical administrative tasks. The lesser the job-stress and feeling of "burn out", the higher the job satisfaction and general well-being in a typical healthcare setting (Young et al, 2008). It is likely that the lesser number of alarms in modified setting allowed participants to complete all tasks without time pressure and be engaged with the system, which was reflected in higher satisfaction score. The only difference between default and modified experimental set up(s) was the total number of alarms. Therefore, changes observed in care provider experience and overall satisfaction were most likely associated with modifications in alarm related workload. A larger sample population and other types of monitoring devices are needed to determine if alarm workload is the causal factor.

Hypothesis 1: Modified alarm setting will have lower workload index than default alarm setting.

This hypothesis was supported. Comparative analysis, using independent samples t-test, showed that the perceived workload, reported using NASA-TLX, was significantly different between two alarm settings. As mentioned previously, participants reported higher workload under default settings than modified alarm settings. Higher mental and temporal demands encountered by participants in default alarm settings contributed to the substantial difference observed between the two alarm settings. There are a lot of decisions to be made when the number of alarms is excessive, and these decisions frequently need to be made under considerable time pressure. Thus, the temporal demand was clearly the highest among all subscales for default setting participants. Frustration level was also higher for participants in the default-setting group. As evidenced, different dimensions of workload exist during alarm management, and they are not equivalent to one another.

Hypothesis 2: Higher alarm workload corresponds to a higher number of committed errors during alarm management.

This hypothesis was supported. The results indicate that participants in the default-setting group with a higher workload erred more frequently than their counterparts in the modified setting group. The results should be interpreted in light of reported temporal demands across these two settings. Evidently, the temporal demand was higher in the default-setting group, because the number of alarms requiring attention was excessive and exceeded the participants' capacities. They felt the time pressure as they addressed one alarm after another while completing patient tasks simultaneously. As

shown in the Table 4.2, the average error %, in default setting (mean = 9.47; SD = 5.99), was 3.6 times more than modified setting (mean = 2.61; SD = 4.48). In a typical hospital setting, many devices support each patient, and nurses are responsible for multiple patients. Thus, each nurse has to attend many alarm signals and calls emitted by these devices. Therefore, there are numerous opportunities to commit errors inadvertently under time pressure. Reduction or removal of these opportunities will likely reduce the probability of making an error.

Hypothesis 3: Higher alarm workload environment corresponds to lower response rate. That is, the higher the alarm workload, the fewer the number of alarms that will be addressed.

This hypothesis was supported. The results clearly show that the higher workload in the default setting prevented participants from responding to the higher number of alarms. Higher temporal and mental demands are inherent to high workload tasks. This was not necessarily unexpected. It is logical that more alarms will produce more troubleshooting tasks, time pressure, and physical expenditure. As such, the excessive number of alarms, typical in default setting, will quickly overwhelm nurses and assistants. Nurses were not readily willing to answer an alarm if it occurred during the performance of other patient care tasks such as getting medication from pharmacy, administering medication or performing patient handoff. It is likely that they experienced cognitive shift—a change in focus when switching from one task to another or moving from one patient to another—and allowed alarms to go on. Though it is reasonable to let the insignificant alarms go unanswered, the consequences of unanswered critical alarms

would be dire. Any effort, incremental or substantial, made to increase specificity of clinical alarms and removal of unwanted alarms would yield significant clinical benefits.

Hypothesis 4: Modification of alarm threshold(s) will result in lower mental workload and lower error rate.

This hypothesis was supported. The overall perceived workload index score was lower in modified setting (mean = 52.4; SD = 2.3) than in default setting (57.6; SD = 2.6), a statistically significant difference of 5.22 (95% CI, 3.39 to 7.05), t (38) = 5.84, $p < 0.05$. As previously explained, customization of alarm limits to patient condition contributed to lower mental workload. It is well known that complex and tedious tasks result in higher workload. Removal of non-essential alarms and notifications from the patients' monitors left nurses with alarms that are truly clinically significant and needed to be addressed. As nurses started trusting that the alarms occurring time and again were significant, they attempted to address every alarm presented and ensured they addressed each one accurately. As hypothesized, there was a difference in the number of committed errors between the different alarm threshold settings. Participants under modified alarm threshold settings committed fewer errors than their counterparts did in the default settings group, because the likelihood of committing errors is far lower due to removal of non-essential alarms in modified setting.

Hypothesis 5: Lower mental workload in nurses corresponds to better patient care and overall satisfaction.

This hypothesis was supported. As hypothesized, alarm customization resulted in lower mental workload and provided a better patient care environment for participants. Modified setting received better ratings than default setting from patient care quality and

satisfaction perspectives. Quality of patient care and nurse satisfaction were included as process and performance indications in this study. These indicators focus on the nature and amount of care nurses provided during the hospital stay. Higher scores for these items indicate that the study participants were able to provide quality care and complete other unit-level duties. Researchers have found that only 1- 10% of clinical alarms resulted in a change in care, and the remaining 90-99% were false or nuisance (Konkani et al, 2012; Way et al, 2014). Based on the findings from this study, it is reasonable to conclude that alarm workload is a modifiable work system factor that affects both patient and healthcare provider outcomes, such as engagement and satisfaction with the environment in which they work. Excessive alarms will interrupt care and increase the likelihood of missing a life-saving critical alarm (Funk et al., 2013; Gorges, 2009). Therefore, hospital administrators should prioritize alarms for every patient-unit and develop protocols to monitor ones that are essential for patient safety.

4.15 Conclusion

Nursing is a high workload profession, and excessive workload, particularly alarm related, has been shown to have an adverse effect on nurses' well-being, job satisfaction, patient care, and safety. Complex work such as nursing care in hospitals involves constant attention to primary tasks (e.g. patient assessments, medication administration) and to intermittent secondary tasks, such as responding to alarms. In such dual-task systems, time and effort spent responding to alarms distracts from the primary tasks. As primary task workload increases, alarm task performance typically worsens, particularly when alarm reliability is low. Alarm response rates in such cases may be low because the operator must choose an action based on the relative urgency of the primary and

secondary tasks. Researchers have linked increased workload to stress, burnout, anxiety, and increased turnover in nursing and other healthcare service fields. Workload is objective, involving a specific task to be completed, yet also subjectively based on the perception of the worker. The subjective perspective of workload can be influenced by the work environment, time pressure to complete the task, and the individual's prior experiences. Alarm customization is a frequently recommended solution to reduce alarm fatigue, as it frees up nurses' cognitive abilities so they can pay attention to alarms and address the important ones. The findings from this study suggest that workload be perceived and initially examined holistically from a broad perspective. However, improvements to the systems that deal with workload should focus on the contributors; hospitals can subsequently examine basic items that contribute to each subscale of the NASA-TLX. This study supports utilizing NASA-TLX to assess and address alarm related workload in a progressive care setting.

4.16 Limitations and future work

The findings of this study should be interpreted in light of the following limitations. Although an appropriate sample size was chosen based on a literature review of similar studies in the healthcare field, the sample size was small due to logistical and regulatory constraints (Health Insurance Portability and Accountability Act). The sample size should be increased and include nurses from various areas of the hospital, such as intensive care, cardiac care, neonatal care, and neurology, as each type of care uses different devices. The current study was confined to research in one geographical region – Seattle in the Pacific Northwest area. It is suggested that future endeavors into alarm workload in nursing examine more than one hospital for comparison of data and

generalization. This study used the mostly frequently used physiological monitoring device. Many types of devices typically serve a hospitalized patient, and a variety of alarms is possible from these devices. Therefore, future studies should include other frequently used devices to assess mental workload.

In this study, mental workload was assessed through a subjective measurement method. There are some disadvantages of using subjective measures. Subjective mental workload measures have been shown to be dependent on localized work culture. It is well known that the healthcare field has regional cultures. What is perceived as normal in one area hospital in the Midwest of the U.S may not be considered normal at hospitals on the west coast. Furthermore, a more focused analysis of nurses' emotions could be helpful to generate a more complete profile of their workload. For example, studying the emotional effects of difficult to trace alarms, non-resolvable or intractable alarm conditions, or other complex patient care tasks while managing alarms could offer a more dynamic account of workload drivers in alarm management.

Another disadvantage is the time gap between completion of experimental tasks and completion of the survey. If the assessment is conducted after some time delay, there is a possibility that the participant may fail to remember all the workload experienced and how he or she felt while performing the task. Though appropriate steps were taken to limit this disadvantage, it was still inherent due to cleaning and returning of the props back to their location. Future studies could supplement subjective measures with objective methods such as galvanic skin response, eye blinking, and heart rate variability (HRV). With availability of smart phones and hand-held computing media, measurements could be taken live.

4.17 References

Abelson, J., Li, K., Wilson, G., & Shields, K. (2016). Supporting quality public and patient engagement in health system organizations: development and usability testing of the Public and Patient Engagement Evaluation Tool. *Health Expectations*, 19(4), 817-827.

Billinghurst, F., Morgan, B., & Arthur, H. M. (2003). Patient and nurse-related implications of remote cardiac telemetry. *Clinical Nursing Research*, 12(4), 356–370.

Bogaert, V., Clarke, S., Wouters, K., & Franck, E. (2013). Impacts of unit-level nurse practice environment, workload and burnout on nurse-reported outcomes in psychiatric hospitals: a multilevel modelling approach. *International Journal of Nursing Studies*, 50(3), 357-365.

Brower, E. J. (2017). Origins of Evidence-Based Practice and What it means for Nurses. *International Journal of Childbirth Education*, 32(2), 14-18.

Brown, S. L., Bogner, M.S., Parmentier, C. M., & Taylor, J. B. (1997). Human Error and Patient-Controlled Analgesia Pumps. *Journal of Intravenous Nursing*, 20, 311-316.

Buist, M. D., Bernard, S., Nguyen, T. V., & Moore, G., et.al. (2004). Association between clinically abnormal observations and subsequent in-hospital mortality: a prospective study. *Resuscitation*, 62 (2), 137–141

Burgess, L. P., Herdman, T. H., Berg, B. W., & Feaster, W. W. et al. (2009). Alarm limit settings for early warning systems to identify at-risk patients. *Journal of Advanced Nursing*, 65(9), 1844–1852.

Carayon, P., & Gurses, A. P. (2005). A human factors engineering conceptual framework of nursing workload and patient safety in intensive care units. *Intensive and Critical Care Nursing*, 21, 284-301.

Chopra, V., & McMahon, L. F, Jr. (2014). Redesigning hospital alarms for patient safety: alarmed and potentially dangerous. *Journal of American Medical Association*, 311(12), 1199–1202.

Cvach, M., Frank, R., Doyle, P., & Stevens, Z. (2013). Use of pagers with an alarm escalation system to reduce cardiac monitor alarm signals. *Journal of Nursing Care Quality*, 97, 1-9.

Dandoy, C E., Davies, S. M., & Flesch, L., et al. (2014). A team-based approach to reducing cardiac monitor alarms. *Pediatrics*, 134(6), 1686–1694.

Deb, S., & Claudio, D. (2017). Alarm fatigue and its influence on staff performance. *IIE Transactions on Healthcare Systems Engineering,* 5, 183–196.

Dressler, R., Dryer, M. M., Coletti, C., & Mahoney, D. et al. (2014). Altering overuse of cardiac telemetry in non-intensive care unit settings by hardwiring the use of American Heart Association guidelines. *Journal of American Medical Association Internal Medicine*, 174(11), 1852–1854.

Drew, B. J., & Funk, M. (2006). Practice standards for ECG monitoring in hospital settings: executive summary and guide for implementation. *Critical Care Nursing clinics of North America,* 18(2), 157–168.

Dye, E., & Wells, N. (2017). Subjective and objective measurement of neonatal nurse practitioner workload. *The National Association of Neonatal Nurses*, 17(4), 3-12.

Ead, H. (2015). Change fatigue in healthcare professionals – An issue of workload or human factors engineering?. *Journal of Perianesthesia Nursing*, 30(6), 504-515.

ECRI Institute. Top 10 health technology hazards for 2013. Health Devices. 2014, 41, 342–365.Retrieved from: https://www.ecri.org/Forms/Pages/ECRI-Institute-2013-Top-10-Hazards.aspx. Last accessed: Nov 17, 2017.

Feder, S., & Funk, M. (2013). Over-monitoring and alarm fatigue: For who do the bells toll?. *Heart Lung*, 42(6), 395-396.

Funk, M., Stephens K., May, J., & Fennie, K., et al. (2013). An alarming rate of unnecessary monitoring in the Practical Use of the Latest Standards of Electrocardiography (PULSE) trial. *Journal of American College Cardiology*, 61, 1496-1498.

Gazarian, P. K. (2014). Nurses' response to frequency and types of electrocardiography alarms in a noncritical care setting: A descriptive study. *International Journal of Nursing Studies*, 51(2), 190–197.

Gomez, T., Sowan, A. K., Tarriela, A., & Reed C., et al. (2015). Reducing non-actionable alarms of cardiac monitors as a strategy to improve alarm safety in a transplant/cardiac ICU. *Critical Care Nurse*, 35(2), 72–73.

Görges, M. M. (2009). Improving alarm performance in the medical intensive care unit using delays and clinical context. *Anesthesia and Analgesia*, 108, 1546–1552.

Gross, B., Dahl, D., & Nielson, L. (2011). Physiologic monitoring alarm load on medical/surgical floors of a community hospital. *Horizons* (Spring Suppl.), 29–36.

Grossman, S.A., Shapiro, N. I., Mottley, J. L., & Sanchez, L., et al. (2011). Is telemetry useful in evaluating chest pain patients in an observation unit? *Internal Emergency Medicine*, 6, 543–546.

Harris, R. M., Manavizadeh, J., McPherson, D. J., & Smith, L. (2011). Do you hear Bells? The increasing problem of alarm fatigue. *Pennsylvania Nurse*, 66(1), 10–13.

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX) 20 years later. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50(9), 55-68.

Hart. S. G., & Staveland, L.F.E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati, (Eds.), Human Mental Workload (pp. 139-183). Amsterdam: North Holland.

Holden, R. J., Patel, N. R., Scanlon, M. C., & Shalaby, T.M., et al. (2010). Effects of mental demands during dispensing on perceived medication safety and employee well-being: a study of workload in pediatric hospital pharmacies. *Research in Social and Administrative Pharmacy*, 6(4), 293-306.

Hoonakker, P., Carayon P., & Gurses, A. (2011). Measuring workload of ICU nurses with a questionnaire survey: the NASA Task Load Index (TLX). *IIE Transactions on Health Systems Engineering*, 2011, 1, 131-143.

Hravnak, M., DeVita, M.A., Clontz, A., & Edwards, L. (2011). Cardiorespiratory instability before and after implementing an integrated monitoring system. *Critical Care Medicine*, 39(1) 65–72.

Hyde, S., Fulbrook, P., Fenton, K., & Kilshaw, M. (2009). A clinical improvement project to develop and implement a decision-making framework for the use of seclusion. *International Journal of Mental Health Nursing*, 18, 398–408.

iPhone App; https://humansystems.arc.nasa.gov/groups/tlx/tlxapp.php. Last accessed: Aug, 2017.

Johnson, R. K., Hagadorn, I. J., & Sink, W. D. (2017). Alarm safety and alarm fatigue. *Clinical Perinatology*, 44, 713-728.

Kaminski, M. M., Meier, S., & Staebler, S. (2015). National association of neonatal nurse practitioners workforce survey. *Advanced Neonatal Care*, 15, 182-190.

Kanwar, M., Fare, R., Minnick, S., & Rosman, H., et al. (2008). Inpatient cardiac telemetry monitoring: are we overdoing it? *Journal of Clinical Outcomes Management*, 15(1), 16–20.

Konkani, A., Oakley, B., & Bauld, T. J. (2012). Reducing hospital noise: A review of medical device alarm management. *Biomedical Instrumentation & Technology*, 46(6), 478–487.

Liu, Y., & Pecht, M. G. (2011). Reduction of motion artifacts in electrocardiogram monitoring using an optical sensor. *Biomedical Instrumentation & Technology*, 2, 155–163.

Malacrida, R., Bomio, D., Matathia, R., & Suter, M. P. (1991). Computer-aided self-observation psychological stressors in an ICU. *International Journal of Clinical Monitoring and Computing,* 8(3), 201-205.

Melynk, B. M., & Fineout-Overholt, E. (2015). Evidence-based practice in nursing & healthcare: A guide to best practice (3rd ed.) (pp. 18). Philadelphia, PA: Wolters Kluwer Health.

Mohammadi, M., Mazloumi, A., Kazemi, Z., & Zeraati, H. (2016). Evaluation of mental workload among ICU ward's nurses. *Health Promotion Perspectives*, 5(4), 280-287.

Munten, G., van den Bogaard, J., Cox, K., & Garretsen, H. et al. (2010). Implementation of evidence-based practice in nursing using action research: a review. *Worldviews Evidence based nursing*, 7(3), 135-157.

Myny, D., Goubergen, D., Gobert, M., & Vanderwee, K., et al. (2011). Non-direct patient care factors influencing nursing workload: a review of the literature. *Journal of Advanced Nursing,* 67(10), 2109-29.

Neill, D. (2012). Nursing workload and the changing healthcare environment: A review of the literature. *Education, Practice and Research*, 1(2), 132-143.

Pergher, A. K., & Silva, R. C. (2013). Alarms fatigue: Integrative review. *Journal of Nursing*, 7(7), 1241–1245.

Physician-Patient Alliance for Health and Safety (2013). 5 resources on the costs of adverse events. Retrieved from http://www.ppahs.org/2013/02/5-resources-on-the-costs-ofadverse-events. Last accessed: Oct 19, 2017.

Rosman, E. C., Blaufox, A. D., Menco. A., & Trope. R. et al. (2013). What are we missing? Arrhythmia detection in the pediatric intensive care unit. *Journal of Pediatrics*, 163(2), 511–514.

Sachdev, M., Feltics, B., Lai, S., & Dalal, D., et.al. (2010). Failure in short-term prediction of ventricular tachycardia and ventricular fibrillation from continuous

ECG in intensive care unit patients. *Journal of Electrocardiology*, 43(5), 400–407.

Sackett, D. L., Straus S. E., Richardson W. S., & Haynes, R. (2000) Evidence based medicine. How to practice and teach EBM, 2nd ed. Churchill Livingstone, UK: Edinburgh, pp 73-141.

Saunders, H., & Vehviläinen-Julkunen, K. (2016). The state of readiness for evidence-based practice among nurses: An integrative review. *International Journal of Nursing Studies*, 56, 128-140.

Soh, T, & Crumpton, L.L. (1996). An evaluation of fatigue indicators while performing nursing tasks. Paper presented at the Industrial Engineering Research, Norcross, GA, pp.130-168.

Srinivasa, E., Mankoo, J., & Kerr, C. (2017). An evidence-based approach to reducing cardiac telemetry alarm fatigue. *Worldviews Evidence Based Nursing*, 14(4), 265-273.

Stevens, R. K. (2013). The impact of evidence-based practice in nursing and the next big ideas. *American Nursing Association*, 18(20), 33-41.

Szczurak, T., Kaminska, B., & Szpak, A. (2007). Estimation of the psychological load in the performance of nurses' work based on subjective fatigue symptoms. *Advances in Medical Sciences*, 52(1), 102-104.

Taenzer, A. H., Pyke, J. B., McGrath, S. P., & Blike, G.T. (2010). Impact of pulse oximetry surveillance on rescue events and intensive care unit transfers: a before-and-after concurrence study. *Anesthesiology*, 112(2), 282–287.

The Joint Commission. (2013). Sentinel Event Alert about FDA report. https://www.jointcommission.org/assets/1/18/SEA_50_alarms_4_5_13_FINAL1.PDF Last accessed: Oct 19, 2017.

Tsang, P.S., & Vidulich, M.A. (2006) Mental Workload and Situation Awareness. In: Salvendy, G., Ed., Handbook of Human Factors and Ergonomics, 243-268.

Tubbs-Cooley H. L., Pickler R.H., & Meinzen-Derr, J. K. (2014). Missed oral feeding opportunities and preterm infants' time to achieve full oral feedings and neonatal intensive care unit discharge. *American Journal of Perinatology*, 32(1), 1-8.

Turmell, J. W., Coke, L., Catinella, R., & Hosford, T. (2017). Alarm Fatigue: Use of an evidence-based alarm management strategy. *Journal of Nursing Care Quality*, 32(1), 47-54.

Varpio, L., Kuziemsky, C., Macdonald, C., & King, W. J. (2012). The helpful or hindering effects of in-hospital patient monitor alarms on nurses: A qualitative analysis. *Computer, Informatics, and Nursing*, 30(4), 210–217.

Veltman, J. A. (2002). A comparative study of psychophysiological reactions during simulator and real flight. *The International Journal of Aviation Psychology*, 12(1), 33-48.

Way, R. B., Beer, S. A., & Wilson, S. J. (2014). Whats that noise? Bedside monitoring in the Emergency Department. *International Emergency Nursing*, 22(4):197–201.

Welch, J. (2011). An evidence-based approach to reduce nuisance alarms and alarm fatigue. Biomedical Instrumentation & Technology: *Alarm Systems*, 45(1), 46–52.

Weng, Y. H., Chen, C., Kuo, K. N., & Yang, C. Y. (2015). Implementation of evidence-based practice in relation to a clinical nursing ladder system: a national survey in Taiwan. *Worldviews Evidence Based Nursing*, 12(1), 22-30.

Wolf, L. D., Potter, P., Sledge, J. A., & Boxerman, S. B. (2006). Describing nurses' work: Combining quantitative and qualitative analysis. *Human Factors*, 48(1), 5- 14.

Yamase, H. (2003). Development of a comprehensive scoring system to measure multifaceted nursing workloads in ICU. *Nursing and Health Sciences*, 5, 299-308.

Young, G., Zavelina, L., & Hooper, V. (2008). Assessment of workload using NASA Task Load Index in perianesthesia nursing. *American society for perianesthesia nurses*, 08, 2302-2305.

Zboril-Benson, L.R. (2002). Why nurses are calling in sick: The impact of health-care restructuring. *Canadian Journal of Nursing Research*, 33(4), 89-107.

CHAPTER V

CONCLUSION

5.1 Introduction

The number of alarm-equipped medical devices used to assist patient care is rising with the surge in demand for high quality care, contributing to exponential growth in the use of alarm systems. These devices monitor patients and administer medication with minimal human intervention. Alarms help improve patient safety by serving as early warnings for clinicians. Alarm fatigue occurs when the sheer number of alarms from medical devices overwhelms nurses and their assistants. Alarm safety is well established as one of healthcare's deeply complex and intractable problems. This can result in desensitization to the alarms, which can lead to incorrectly responding or altogether missing them. Thus, alarm fatigue defeats the purposes of designing alarms within the device. Alarm fatigue also influences patient satisfaction measures when it affects patient sleep and anxiety, as well as that of family members or caregivers accompanying the patient. Though several initiatives have been taken by various government and non-profit entities, improvements appear to be minimal based on recent adverse event filing with the Food and Drug Administration (FDA-MAUDE) on alarm related sentinel events. The FDA's MAUDE data clearly shows that the individual hospitals need to take a systematic and interdisciplinary approach to alarm safety issues. Focus on alarm types,

standardization of alarm management practices, improvement of staff knowledge, and customization of alarm settings should be part of that effort. Though technology can aid the nurses and their assistants in managing alarm related hazards, it alone cannot solve the problem of alarm fatigue. Modified training and alarm settings, when used in conjunction with other frequently recommended easy-to-adopt solutions, such as standardization, smart phone based applications, and alarm escalation technologies, can help reduce the problem. This approach to alarm safety has the potential to prevent alarm fatigue among hospital staff and reduce the risk for patient-related sentinel events.

5.2 Impact of generalized and individualized feedback on metacognitive prediction accuracy

Calibration of performance is crucial because it allows any learner to engage in appropriate comprehension monitoring and later apply learned concepts when there is a need. This study investigated the effects of training method and feedback provided on metacognitive calibration over the course of three exams. Participants' metacognitive prediction accuracy (measured by calibration) and exam performance were assessed after two types of training interventions – classroom and one-to-one training. High-level feedback about calibration and classroom performance was provided to classroom participants, whereas individualized feedback on incorrect answers and ways to improve metacognitive prediction accuracy was provided to one-to-one participants. Study participants started the study overconfident in their predictions and ended up under-confident, irrespective of their training method.

The findings from this study show that generic feedback provided to classroom participants helps improve metacognitive feedback. The findings also reveal that

participants' prediction accuracy was slightly better than their one-to-one peers and improved overtime. Thus, the generic feedback provided acted as performance feedback for classroom training. On the other hand, individualized feedback provided to one-to-one participants discussed incorrect answers, provided right answers with reasoning, and discussed ways to improve metacognitive prediction so they can be better calibrated in the future. The surprising finding was that prediction accuracy did not improve over time for one-to-one participants. Furthermore, they were poorly calibrated compared to their classroom trained peers on exams 2 and 3. Results show that the individualized feedback provided was likely environmental in nature and did not improved metacognitive prediction accuracy. Due to excess knowledge gained through individualized feedback and one-to-one attention, participants gave their predictions based on what they did not know rather than what they knew, whereas the classroom participants' predictions were based on what they knew from classroom training. This difference explains the wide gap in their calibration scores, as medical device alarms is a knowledge ocean.

Although one-to-one trained participants' prediction accuracy did not improve, their content knowledge increased significantly. More importantly, they were able to identify their weaknesses and focused their attention towards improving them, resulting in better scores than their testing counterparts. In fact, many achieved near-perfect scores on Exam 3. The vast difference in exam scores between one-to-one participants and the rest shows that the former likely altered their study habits based on the feedback. The findings from this study suggest that a combination of performance and environmental feedback is necessary for optimal clinical performance.

5.3 Influence of one-on-one training on alarm management

To investigate the theory that training methods influence how nurses and their assistants grasp alarm related knowledge, this study examined the effects of training procedures on classroom performance through assessment exams post-training. The findings show that training can affect alarm knowledge when users receive modified training in a one-to-one fashion. Results show that one-to-one training improved nurses' and their assistants' alarm knowledge substantially. This is true irrespective of the device complexity.

The results also confirm that existing training methods used by biomedical departments do not adequately provide the knowledge that nurses and their assistants require. It is possible that existing training methods are application oriented and limit their scope to troubleshooting techniques of various alarms, as the intent is to quickly resolve the alarm and move to other tasks. However, the results of this study show that training methods that include working principles and some theory would improve nurses' understanding of medical device alarms and increase their knowledge. Such training, when customized based on nurses' technical grasp, will benefit not only nurses and patients, but also the entire healthcare system. Any increase in knowledge would also likely increase nurses' self-confidence in handling complex alarm related clinical situations.

Today's complex healthcare environment pressures nurses and their aides to become fast leaner(s) and cost-conscious practitioners while working with fewer resources. Because assimilation of study materials and new concepts vary from person-to-person, it is impossible for everyone to become a quick study. Therefore, the training

methods used to educate healthcare staff on complex concepts, such as alarm management, need to be customized based on one's competence and speed of grasping fundamentals. There are two powerful benefits to this approach: (i) the training examples used in class would pertain directly to the individual employee's circumstance and experience. The employee could, therefore, understand how the examples and situations in class relate to his/her job and connect with the examples if such situations arise again, and (ii) two-way interaction between the trainer and trainee gets the employee more engaged, and the professional relationship established through this interaction will allow trainees to reach out to the trainer later when there is a complex situation. In addition, customization will also result in increased knowledge retention. Moreover, trainers can prepare the training material better when they know their audience's areas of strengths and weaknesses and ways in which they most effectively absorb training material. Across the United States, hospitals' biomedical engineering departments usually act as trainers of employees. Trainees, when acquainted better with trainers, will be at ease to discuss and ask for help if there is a need. By customizing the training delivered to nurses, hospitals can reap the benefits of improved enthusiasm and acceptance of training, which is crucial to implement new regulations and policies. Over time, customized training will likely lead to increased employee competence and independent learning that is less reliant on biomedical departments.

5.4 Effects of modified alarm threshold limits on alarm workload, response and error rates, and patient car experience

The quality of patient care during a hospital visit is essentially determined by the quality of devices and technology used for treatment, quality of training, competence of

health care providers, and efficiency of support systems. Unfortunately, the devices that are intended to improve the quality of healthcare and produce positive outcomes through accurate physiological monitoring actually harm the patient and adversely affect the quality due to the exorbitance in one of its features. To investigate the theory that modifying the alarm threshold limits based on patient condition would positively affect patient care and reduce workload, this study examined the effect of modified alarm threshold limits on workload and performance measures, such as alarm response rate and number of errors committed, along with care provider experience. The study measured the workload under different alarm settings – default and modified. The findings show that removal of certain non-essential alarms, based on patient condition, can result in better care provider experience, reduced mental workload, and higher overall satisfaction.

In addition, alarm customization allows nurses to respond to a greater number of alarms accurately than they would with a default alarm setting, which may contribute to an increase in response errors, possibly due to time-pressure. That is, reducing the number of alarms allowed nurses to respond to almost all of the remaining alarms accurately. Though a few studies have shown improvements in alarm fatigue through alarm customization and false alarm reduction, previously published studies have not established the relationships between alarm customization, nurses' workload, and performance measures, such as alarm response rate and error rate. The results from this study clearly show that the number of managed alarms is directly proportional to workload and the number of errors (error rate) committed and inversely proportional to alarm response rate and care provider experience. Furthermore, most of the previous studies in alarm customization adopted customization at the unit level, where every

patient moved to the unit had the same monitor settings, whereas, this study modified at the patient level. The results from this study affirm one of the frequently recommended solutions by alarm fatigue researchers – alarm customization can reduce alarm related workload and positively influence alarm management experience. The findings from this study open the scope for further research to reduce alarm fatigue in nurses by analyzing the influences of all possible factors that can cause alarm fatigue. The scope of the current study was limited to a progressive step-down unit where the incoming patients were relatively stable. Therefore, additional studies in other environments, such as intensive care and cardiac care, and extensive real-world data are needed before generalization and implementation.

5.5 Summary

Alarm safety, one of the National Patient Safety Goals in 2017, is a difficult problem to solve. This study assessed the effectiveness of two frequently recommended solutions to reduce clinical alarm fatigue. This study also investigated whether altering the training methods currently used to train nurses in medical device alarm management and whether customizing alarm thresholds will better equip the nurses for managing clinical alarms. Training nurses on metacognitive components will help them in appropriately responding to clinical situations with an appropriate level of confidence. Results show that training and feedback, when individualized in a one-to-one setting, substantially improve study participants' domain specific knowledge. However, calibration and prediction accuracy does not improve with training and feedback over time. Trained participants were poorly calibrated, on the under-confidence side, compared to their untrained peers. Metacognition appears to be somewhat resistant to

improvements even after multiple training sessions, because trained professionals like nurses tend to predict their performance based on what they do not know rather than what they know. Furthermore, they do not want to come across as the presumptuous "know-it-all" person in their peer group. This is perfectly acceptable in high impact fields, such as healthcare, because clinicians typically err on the conservative side or are reluctant to test the unknowns and learn through 'trial and error'.

This study evaluated the impact of adjusting alarm threshold limits on physiological monitoring, based on individual patients' clinical conditions, on clinicians' performance attributes – alarm response rate and number of errors committed. In addition, the study also measured the subjective workload for two alarm settings. Removal of unnecessary alarms based on patient condition resulted in a lower number of alarms in a modified setting than in the default setting, where the manufacturer-set alarms are used for physiological monitoring. Consequently, nurses and their aides responded better to the remaining alarms. Hence, the number of errors committed was relatively lower in modified setting than the default setting. Evidence for optimal alarm settings for physiological monitors and cardiac devices are abundant. Hospital administrators should make every effort to develop appropriate threshold levels for various physiological measures clinicians monitor for typical patient conditions. This will help reduce the alarm burden for nurses and their aides significantly. In addition, there are other benefits in customizing alarm threshold limits – lower workload, better care provider experience, and increased overall satisfaction. Additional studies across other types of devices are needed to generalize the benefits of alarm limit customization. Addressing alarm fatigue

requires regulatory bodies, device manufacturers, and hospital administrators recognizing the importance of training and the amount of alarm related workload.

The results from this research confirm that the two most frequently recommended solutions — better training and alarm customization — will help improve alarm system safety significantly. As is well established, alarm management is very complex and as such requires other systemic improvements in layout, protocols, work stream, and standard operating processes. Clinical alarm management is in nascent stages in many hospitals across the U.S. This is the best time window to implement policy changes, develop new training methods, and process modifications.

APPENDIX A

IRB PACKET

A.1 Demographic questionnaire

Please mark your selection clearly with a 'x' or '✓'

1. Are you currently practicing as a registered nurse (RN)?

   Yes [ ] No [ ]

2. How many years of experience do you have in managing medical device alarms?
   None [ ]
   Less than 1 year [ ]
   1-3 Years [ ]
   3-5 Years [ ]
   More than 5 Years [ ]

3. Have you received any training on medical device alarms?
   Yes [ ] No [ ]

4. Do you feel the training provided by your institution is adequate?
   Yes [ ] No [ ]

5. Please provide any comments if you have chosen 'No' for Q4 & Q5

6. Please choose all applicable items from these options:
   I have an Associate's Degree in my field [ ]
   I have a Bachelor's Degree in my field [ ]
   I have a graduate degree (M.S, Ph.D, DNP etc) in my field [ ]
   I have cleared NCLEX or other nursing licensing exam [ ]

7. Did your assigned unit provide any training on alarm management?

   Yes [ ] No [ ]

8. Do you have any certifications on alarm safety? (include all internal, external, manufacturer certified etc)

A.2 Care provider experience & overall satisfaction

Please mark with an 'x' or '✓'

1. How satisfied are you with the level of care provided to all your patients (mannequins)?
   [1] Very Dissatisfied
   [2] Dissatisfied
   [3] Neutral; Neither Satisfied nor Dissatisfied
   [4] Satisfied
   [5] Very Satisfied

2. If **NOT SATISFIED**, what prevented you from administering care? Choose all that apply.
   Training issue [ ]
   Too many tasks [ ]
   Too many alarms [ ]
   Not a fit for this floor [ ]
   Other reasons [ ]

3. Please provide any comments that could provide more clarity or support your choices for Q1 and Q2.

4. Overall, how satisfied are you with number of alarms you needed to address while providing patient care?

   [1] Very Dissatisfied

   [2] Dissatisfied

   [3] Neutral; Neither Satisfied nor Dissatisfied

   [4] Satisfied

   [5] Very Satisfied

5. How do you rate number of alarms presented?

[ ] Far Too Few [ ] Too Few [ ] About Right [ ] Too Many [ ] Far Too Many

A.3 Exam questions

MX-40 ®

Section 1 [Device related questions]

1) Device can be sterilized using: …… (a) Hydrogen Peroxide (b) Eto (c) Hydrogen sulfide (d) Benzoyl Peroxide
2) Max number of screen format: ……. (a) 5 (b) 6 (c) 7 (d) 8
3) How many channels of real time wave forms are available? (a) 2 (b) 3 (c) 4(d) 5
4) The technology used by MX-40 for measurements: (a) Short range radio (b) RFID (c) Medical implant(d) Common communications
5) Intellivue MX40 needs to be charged for at least …. Hours? (a) 3(b) 6 (c) 9(d)12
6) Device can be wirelessly connected to ….(a) all Philips devices (b) only Intellivue devices (c) all monitors through Bluetooth (d) cannot be connected
7) Screen format is programmable based on hospital floor procedures (a) Yes, possible (b) No, not possible
8) All alarms sounds are configurable (a) Yes (b) No (c) Depends on the type
9) Who can change batteries? (a) BMET (b) Floor supervisor (3) Manufacturer. Answer (a) 1 only (b) 2 only (c) 3 only (d) All three
10) Intellivue is a firewall-enabled system. Therefore, internet can be accessed. (a) Yes (b) No

Section 2 [Alarm related questions]

11) All alarm for 'leads-off' require action (a) yes (b) no (c) depends
12) Voltage change alarm requires (a) Yes (b) no (c) depends
13) Default setting for low heart rate (a) 70 (b) 75 (c) 72 (d) 60
14) Alarm data is saved for (a) 30 days (b) 60 days (c) 90 days (d) 120 days
15) Red arrhythmia alarm requires (a) intervention by MD (b) no action (c) escalation to telemetry (d) minor adjustment of leads
16) Default end tidal $C0_2$ range is (a) 5-6% (b) 6-7% (c) 7-8% (d) 8-9%
17) Premature ventricular contraction rate is a benign condition (a) Yes (b) No (c) depends
18) Delay range for non-actionable alerts is (a) 1-5 mins (b) depends on unit policy (c) 5-10 minutes (d) depends on hospital protocol
19) All alarms can be suspended for (a) 5 mins (b) depends on unit policy (c) depends on hospital protocol (d) 10 mins
20) Disposable sensors can be used in a high humidity environment for measuring $Sp0_2$ (a) Yes (b) No (c) depends

Section 3 [Troubleshooting]

21) When the sensor position alarms you do (a) follow proper skin prep technique and try placing again (b) Request new sensor (c) mute the alarm (d) Notify MD and floor supervisor
22) When 'lead-off' alarm occurs (a) notify MD and floor supervisor (b) Request new leads and attempt to replace (c) Mute the alarm (d) Unplug the monitor and re-start
23) "Artifact" alarm occurs (a) Suspend the alarm (b) Notify MD and floor supervisor (c) Unplug the monitor and restart (d) Remove the source of artifact and attempt to restart
24) High BP alarm occurs (a) Review patient chart; review threshold limit and take action (b) Notify care provider and floor supervisor (c) Inject a bolus doe of blood pressure reducing medicine (d) Mute the alarm and carry on with your work
25) When 'missed-beat' alarm occurs (a) Review patient chart; review threshold limit and take action (b) Notify care provider and floor supervisor (c) Using an external device, manually check for the beat(d) Mute the alarm and carry on with your work
26) Double PVC alarm occurs (a) Review patient chart; review threshold limit and take action (b) Notify care provider and floor supervisor (c) Adjust patient position and see it stops (d) Mute the alarm and carry on with your work
27) $SP0_2$ low alarm occurs (a) Review patient chart; review threshold limit and take action (b) Notify care provider and floor supervisor (c) Increase 0_2 rate to patient (d) Mute the alarm and carry on with your work
28) When 'vent' alarm occurs (a) Mute the alarm and carry on with your activity (b) Notify floor supervisor and MD (c) Unplug the monitor from the wall outlet (d) Disconnect all leads and sensors
29) When "E-Tachy" alarm occurs (a) Call code protocol (b) Suspend the alarm (c) Perform CPR (d) Manually check using an alternate equipment
30) When "ABP" alarm occurs (a) (a) Call code protocol (b) Suspend the alarm (c) Give bolus does of BP medicine (d) change the sensor cable and check if alarm resolves.

Kangaroo®

Section 1 [Device related questions]

1) Device needs to be sterilized prior to each use (a) Yes (b) No
2) Calibration is required every year and must be done by the Manufacturer (a) Yes (b) No
3) Preventive maintenance is required every year (a) Yes (b) No (C) Depends on usage
4) Kangaroo system requires special cable to connect with its accessories (a) Yes (b) No
5) Kangaroo system must be connected to an UPS (uninterruptible power supply) at all times (a) Yes (b) No
6) Alarms sounds in Kangaroo systems are configurable (a) Yes (b) No
7) The pump mechanism in Kangaroo system is (a) peristaltic (b) hydraulic (c) doppler technology
8) Heat energy is generated during Kangaroo pump operation (a) Yes (b) No (c) depends
9) Cooling Kangaroo pump, when you feel the pump and surroundings, is optional (a) Yes (b) No
10) Kangaroo pump has wireless communication ability (a) yes (b) No

Section 2 [Alarm related questions]

11) System error requires action (a) Yes, power down (b) No, no action (c) depends
12) Hold error occurs when the pump is inactive (a) Yes, for 10 minutes (b) No (c) Yes, for 30 minutes or more
13) Rotor error appears during (a) Running (b) Priming (c) Both (d) Neither
14) Feed error occurs when the enteral formula is no longer being delivered due to (a) Empty bag (b) Clog (c) Both (d) Neither
15) Feed error occurs when the enteral formula is no longer being delivered due to (a) Empty bag (b) Clog (c) Both (d) Neither
16) Flow error occurs due to a clog between (a) Pump and patient (b) Patient and bag (c) Pump and bag (d) airlock in the system
17) The 'PUMP SET DISLODGED' screen will appear if the black ring retainer (MISTIC) is not properly loaded in the MISTIC pocket in the Pump Set loading area. (a) Yes (b) No (c) depend
18) The 'Pump Set' usage warning indicator will blink on the RUNNING screen if a Pump Set has been used for … or more hours (a) 6 (b) 12 (c) 18 (d) 24

19) The 'BATTERY LOW' screen appears and the alarm beeps continuously when the battery needs to be recharged. And you have ….minutes to change (a) 15 (b) 30 (c) 45
20) The 'FEEDING COMPLETE' information screen appears after completion of the programmed feeding (a) True (b) False

Section 3 [Trouble shooting questions]

21) When 'HOLD' error occurs, the pump can be set to run immediately, or the pump can be set to run in a specified number of minutes.(a) Yes (b) No
22) The 'ROTOR ERROR' problem can be resolved by (a) changing out pump set tubing (b) changing out rotor (c) changing out rotor casing
23) The 'FEED ERROR' can be resolved by (a) clearing occlusion (b) increasing feeding rate (c) decreasing feeding rate
24) One way of clearing a detected occlusion is (a) load a new pump set (b) flush (c) aspirate/suction
25) If 'FLUSH ERROR' is detected, then (a) load a new pump set (b) flush (c) aspirate/suction
26) 'FLOW ERROR' is caused by dampness in (a) valve pocket (b) rotor pocket (c) both a and b (d) none of the above
27) 'FLOW ERROR' is also caused by (a) dirt (b) pinched valves (c) damaged rotor (d) None of the above
28) The 'Pump Set Dislodged Error' can be resolved by appropriately positioning the MISTIC retainer (a) True (b) False
29) The Pump Set usage warning indicator will blink on the ……screen (a) Home (b) Running (c) Both Home and Running screens
30) The 'Pump Set >24 hours' warning requires action from the user (a) True (b) False

Alaris 8015

Section 1 [Device related questions]

1) The Alaris unit can be operated manually or in concert with the information exchanged with Alaris Systems Manager (a) True (b) False
2) If communication with the wireless network is interrupted (for example, out of range), the Alaris System can be used, as intended, in the manual mode (a) True (b) False (c) Depends, only on emergency situations
3) Alaris pump model needs wireless network card for operation (a) Yes (b) No (c) Depends on usage and scope

4) The combined use of the Alaris System and Alaris Systems Manager is integrated into a facility's existing network infrastructure with minimal modification (a) True (b) False
5) The Alaris System is designed to operate a maximum of ….. infusion or monitoring modules (a) 4 (b) 6 (c) 8 (d) 10
6) Application of adhesive tape or other materials to the sides of pump unit and modules is an acceptable "work around" technique (a) Yes (b) No (c) Depends on situation
7) By default infusion parameters are set to clear after every (a) 8 hours (b) 24 hours (c) 96 hours (d) 200 hours
8) It is acceptable to use Alaris system near MRI system (a) Yes (b) No (c) Depends; it is acceptable at low Tesla (1.5)
9) You need Manufacturer's support to change from Factory default setting to hospital defined settings (a) True (b) False
10) Connectors on the pumps must be replaced when (i) Blue deposits are found (ii) Green deposits are found (iii) Surface contaminants are found (a) I only (b) II only (c) I, II and III (d) III only

Section 2 [Alarm related questions]

11) 'Channel disconnected' alarm means (i) Module disconnected while in operation (ii) Have a communication problem (a) I only (b) II only (c) I and II (d) Neither
12) Accumulated air-in-line alarm means (i) A large number of air bubbles smaller than current air-in-line limit (ii) Voids and continuity issue (a) I only (b) II only (c) I and II (d) Neither
13) Air-in-line alarm impacts infusion of fluids (a) Yes (b) No
14) Check IV set alarm means administration set is not properly installed (a) Yes (b) No
15) Close door alarm means module door open during an infusion (a) Yes (b) No
16) Occluded-Fluid side/empty container alarm means (i) Upstream occlusion or empty container (ii) Holes of device occluded (a) I only (b) II only (c) I and II (d) Neither
17) Partial occlusion-patient side means (i) Partial occlusion of patient side of IV line detected (ii) Patient's arteries/veins have blockages (a) I only (b) II only (c) I and II (d) Neither
18) 'Pump chamber blocked' alarm means (i) Tubing blocked in pump (ii) Tubing blocked in extension set (a) I only (b) II only (c) I and II (d) Neither
19) Restart channel means module was paused for 2 minutes (a) Yes (b) No
20) Panel unlocked alarm means (i) Tamper resist feature deactivated (ii) Key panel is active (a) I only (b) II only (c) I and II (d) Neither

Section 3 [Trouble Shooting questions]

21) For 'Attach Dose Request Cord' alarm, appropriate action(s) would be (i) Reattach dose request cord (ii) press 'RESTART' key (a) I only (b) II only (c) I and II (d) Neither
22) For 'Channel Disconnected' alarm, an appropriate action would to reattach module and click into place (a) Yes (b) No
23) 'Drive not engaged' alarm can be resolved by (i) opening and closing plunger gripper (ii) adding a driver (a) I only (b) II only (c) I and II (d) Neither
24) 'Channel Error' can be rectified by replacing the module (a) Yes (b) No
25) 'Incorrect concentration' can be rectified by reprogramming (a) Yes (b) No
26) 'Infusion Complete-KVO' alarm should be responded by (i) Unplugging the infusion set (ii) Pressing channel off (a) I only (b) II only (c) I and II (d) Neither
27) 'Syringe Empty' can be resolved by (i) verifying that syringe plunger movement is unimpeded (ii) Verifying that appropriate pressure sensing disc is in use (a) I only (b) II only (c) I and II (d) Neither
28) 'Occluded-Fluid Side' can be resolved by (i) Clearing occlusion on fluid side of instrument (ii) refilling drip chamber (a) I only (b) II only (c) I and II (d) Neither
29) 'Syringe calibration required' can be resolved by replacing module (a) Yes (b) No
30) 'Panel locked' alarm can be rectified by deactivating tamper resist feature using 'Tamper Resist Control on back of PC unit' (a) Yes (b) No

A.4 NASA TLX

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

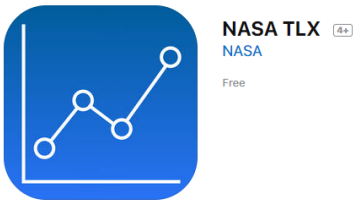

Figure A.1 Schematic of NASA TLX in iPhone® App store

A.5 Informed Consent

Mississippi State University
Data Collection Informed Consent Form for Participation in Research

Title of Research Study: Individualized training and individualized alarm thresholds: Two ways to reduce potential hazardous clinical alarm related incidents

Study Site: Seattle University College of Nursing Simulator Lab, C3108, Seattle WA And Pacific Labs (an external lab) where they have simulator and resources.

Researchers: Mani Shanmugham, Mississippi State University, Dr. Lesley Strawderman, Mississippi State University, Dr. Kari Babski-Reeves, Mississippi State University, Dr.Deborah Eakin, Mississippi State University and Dr. Linkan Bian, Mississippi State University

Purpose: The purpose of this project is to assess (i) whether providing customized training or feedback to nurses will influence two aspects of metacognition – monitoring judgement and control and impact alarm management (ii) assess whether customizing alarm settings to patient conditions will impact nurses' alarm response rates and error rates.

Procedures

Study # 1
Participants will be asked to enroll in 3-week training program on medical devices, alarms and alarm management. Upon recruitment, they will be assigned to one of three groups. Participants are required to attend classes that will be conducted Monday-Friday for an hour (for some participants) or may be more (depending on their assigned group). To keep number of enrolled hours consistent across groups, a minimum of 5 hours attendance is required. The training will take place in a conference room setting or in huddle room setting. The hospital biomedical equipment technician or the principal researcher will teach topics such as working principle, troubleshooting and managing alarms. Three exams will be administered over the course of this training program at various time points (end of each week). Entire experiment will take in a lab simulator with mannequins, models and actual medical device.

Exams will be objective type with a maximum score of 50. Both exams will be of 30 minute duration. Some study participants will receive high-level generic feedback after Exam 1 and prior to starting week 2 class. The feedback will consist of presenting the

mean score and standard deviation (class) for each section, and showing the availability of study materials on various channels (websites, instructor notes, and manufacturer's printed materials). For some participants, answer sheets from Exam 1 will be reviewed individually and weak areas will be identified based on their sectional answers. The experimental session will end with a personalized "Thank You" from the researcher and distribution of financial obligations.

Study # 2

Nurse participants (one at a time) will be asked to check in at one of the nurse bay stations in the simulator. Before starting the session, participants will be provided to an overview of the experiment, and participants will be asked to complete an informed consent form approved by the Mississippi State University IRB and completed a paper demographic questionnaire. Each experimental session will last approximately 30 minutes and each study participant will be presented with few tasks (4 tasks) in the same sequence for all participants. These are simple tasks that a typical Nurse would do every day such as taking a call from pharmacy, checking patient's vitals, receiving a patient from a different department, handoff or discharge a patient etc. The study participants will be asked to repeat the experiment twice (either on a same day or next day). During experimental session, the medical mannequin "M-1" will present alarms to study participants and the researcher will record data (in the observation room) about alarms, their time etc. Participants will be advised to act normally how they would in their everyday work life such as at a typical hospital floor while providing patient care.

Upon completion of experiments, nurse participants will be asked to complete a shot 2-question survey (care provider experience). Participants will be thanked by the researcher and dismissed from the simulator session. Study participants will be requested to go to debriefing room and wait. Someone will request study participants to fill out a NASA-TLX electronic survey on an iPad. This simple survey will take approx. 20-30 minutes to complete.

Risks or Discomforts

There is not more than minimal risk associated with participation in this study.

Benefits

Participants would provide results, which would substantially increase our understanding about metacognition in clinical setting. Results from the study will help the research community in understanding the relationship between individualized training and metacognition. Furthermore, both sides will understand how minimizing the number of alarms impacts care provider satisfaction and patient outcome. Participants will also

benefit from the experience of participating in and exposure to a research project analyzing how to enhance classroom-learning experiences that directly impact medical alarm management success. They will be able to build a body of knowledge on various medical device alarms.

Incentive to participate

Participants who completed exams (in Study #1) will be paid $25. If classes falls during lunch, catered lunch will be provided. No financial compensation has been allocated for study # 2.

Further, number of hours attended will be recorded and handed to participants (in a memo format signed by the principal researcher). They could use this as CEUs (continued education units) and use towards internal certification.

Confidentiality

Individual identities will be protected and will not in any way be connected with any written summary of results that may later be published. Personal information that is collected will be separated from the data collected. Raw data will only be available to the project investigators. Electronic data will be stored on a password-protected computer. All performance data will be stored separately from identifying numbers. Also, please note that these records will be held by a state entity and therefore are subject to disclosure if required by law.

Questions

If you have any questions about this research project, please feel free to contact Mani Shanmugham at 801-673-9973 or Dr. Lesley Strawderman (Faculty Advisor) at 662-325-7214.

For questions regarding your rights as a research participant, or to express concerns or complaints, please feel free to contact the MSU Regulatory Compliance Office by phone at 662-325-3994, by e-mail at irb@research.msstate.edu, or on the web at http://orc.msstate.edu/*participant/*.

In addition to reporting an injury to Mani Shanmugham at 801-673-9973 and to the Regulatory Compliance Office at 662-325-3994, you may be able to obtain limited compensation from the State of Washington if the injury was caused by the negligent act of a state employee where the damage is a result of an act for which payment may be made under §11-46-1, et seq. State Code Annotated 1971. You can also file a claim

through MS State Univ. Please contact the University Police Department at *MSU UNIVERSITY POLICE DEPARTMENT, Williams Building, Mississippi State, MS 39762, (662) 325-2121.*

Voluntary Participation

Please understand that your **participation is voluntary.** Your **refusal to participate will involve no penalty or loss** of benefits to which you are otherwise entitled. You **may discontinue your participation** at any time without penalty or loss of benefits.

Options for Participation

Please initial your choice for the options below:

___The researchers may contact me again to participate in future research activities.

___The researchers may NOT contact me again regarding future research.

Please take all the time you need to read through this document and decide whether you would like to participate in this research study.

If you agree to participate in this research study, please sign below. You will be given a copy of this form for your records.

_____ _____

Participant Signature Date

_____ _____

Investigator Signature Date

MISSISSIPPI STATE UNIVERSITY

Office of Research Compliance

Institutional Review Board for the Protection of Human Subjects in Research
P.O. Box 6223
53 Morgan Avenue
Mississippi State, MS 39762
P. 662.325.3294

www.orc.msstate.edu

NOTICE OF APPROVAL FOR HUMAN RESEARCH

DATE:	August 16, 2017
TO:	Lesley Strawderman, Industrial and Systems Engineering, Deborah Eakin;Kari Reeves;Linkan Bian;Manikantan Shanmugham
FROM:	Jodilyn Roberts, HRPP Officer, MSU HRPP
PROTOCOL TITLE:	Modifying training methods and individualizing alarm thresholds: two ways to reduce potential hazardous clinical alarm related incidents
PROTOCOL NUMBER:	IRB-17-376
	Approval Date: August 16, 2017 Expiration Date: August 16, 2018

This letter is your record of the Human Research Protection Program (HRPP) approval of this study as exempt.

On August 16, 2017, the Mississippi State University Human Research Protection Program approved this study as exempt from federal regulations pertaining to the protection of human research participants. The application qualified for exempt review under CFR 46.101(b)(2).

Exempt studies are subject to the ethical principles articulated in the Belmont Report, found at www.hhs.gov/ohrp/regulations-and-policy/belmont-report/#

If you propose to modify your study, you must receive approval from the HRPP prior to implementing any changes. The HRPP may review the exempt status at that time and request an amendment to your application as non-exempt research.

In order to protect the confidentiality of research participants, we encourage you to destroy private information which can be linked to the identities of individuals as soon as it is reasonable to do so.

The MSU IRB approval for this project will expire on August 16, 2018. If you expect your project to continue beyond this date, you must submit an application for renewal of this HRPP approval. HRPP approval must be maintained for the entire term of your project. Please notify the HRPP when your study is complete. Upon notification, we will close our files pertaining to your study.

If you have any questions relating to the protection of human research participants, please contact the HRPP by phone at 325.3994 or email irb@research.msstate.edu. We wish you success in carrying out your research project.

Figure A.2 Approval letter from MSU IRB

A.6 Script used by nurse educator

1. Introduction of Investigator & Study

Excuse me, sir/ madam OR Name

- **(confirm that you have the correct person if you are contacting a specific patient or potential subject)**

Do you have a minute? My name is __XYZ,
I am a Nurse Educator at Swedish Medical and I am working on a research study with a student from MS State. You received information about this study via recruitment flyer sent via email and on wall posters.

2. Immediate opportunity to opt-out

I am here to follow up on the flyer and to see if you are interested in hearing more about our study. Is it OK for me to continue?

- If individual says "no, not interested" = **stop, say thank you but do not continue.**

If he/she says yes, then continue or make plans to revisit at a more convenient time.

3. Make a BRIEF statement about why he/she was selected/called. Make sure the individual understands that this research is separate from his/her job, clinical care she provides. State like this:

- I am approaching you because we are looking for nurses and their assistants who work with medical and patient support devices equipped with alarms in step down units, progressive care and tertiary care units. This research is totally separate from the care you are providing here and your day job. Whether or not you decide to hear more about the research will not affect your job, patients you care or hospital environment.

4. Ask if he/she is interested in hearing more details.

So, are you interested in hearing some details about the research study?

- If not interested, thank the individual for his/ her time.
- If interested, then move to the consent form and read it.