12-2016

Enhanced Mutual Performance Monitoring to Improve Backup Behaviors and Team Performance

Tiffany Nicole Cooper
Clemson University

Follow this and additional works at: http://tigerprints.clemson.edu/all_theses

Recommended Citation
Cooper, Tiffany Nicole, "Enhanced Mutual Performance Monitoring to Improve Backup Behaviors and Team Performance" (2016). All Theses. 2538.
http://tigerprints.clemson.edu/all_theses/2538

This Thesis is brought to you for free and open access by the Theses at TigerPrints. It has been accepted for inclusion in All Theses by an authorized administrator of TigerPrints. For more information, please contact awesole@clemson.edu.
ENHANCED MUTUAL PERFORMANCE MONITORING TO IMPROVE BACKUP BEHAVIORS AND TEAM PERFORMANCE

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Applied Psychology

by
Tiffany Nicole Cooper
December 2016

Accepted by:
Dr. Fred Switzer, Committee Chair
Dr. Marissa Shuffler
Dr. Peggy Tyler
ABSTRACT

Organizations today are becoming increasingly dependent on teams, as often tasks that need to be completed are too complex for an individual alone. There are a multitude of factors that contribute to how effective a team is; however, an important process that needs to be studied more thoroughly is mutual monitoring and consequential backup behaviors. In the past backup behaviors have been studied solely through a task workload manipulation with methods of mutual monitoring rarely being addressed. The present study explores various types of team monitoring through the use of two experimental conditions: monitoring the task performance levels and a control condition in which no meaningful monitoring is available. This in turn, can be related to team performance and backup behaviors. It was hypothesized that teams that are able to monitor each other’s task levels will have the higher team performance and a higher number of backup behaviors in the presence of legitimate need compared to the control group. Additionally, these differences were explored for differences over time through performance episodes along with examining their physiological compliance (GSR). This study used a chemical plant simulation where teams of three work together. Participants were recruited through Clemson SONA systems for class credit. Findings from this study indicate that the hypotheses were not supported; however there are some interesting conclusions that can be made from this. There are potential implications for face to face vs virtual teams, monitoring assistance perception’s relationship with team emergent states, defining legitimacy of need, and physiological compliance’s potential relationship with intra-team differences.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE PAGE</td>
<td>i</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
</tbody>
</table>

## CHAPTER

1. **INTRODUCTION** ................................................................................................. 1

   Team effectiveness .................................................................................................. 2

2. **MUTUAL MONITORING AND BACKUP BEHAVIORS** .............................................. 5

   Framework ............................................................................................................ 5
   Mutual Performance Monitoring .......................................................................... 8
   Backup Behaviors ................................................................................................. 9

3. **WORKLOAD MEASUREMENT** ............................................................................... 11

   Mental Workload ................................................................................................... 11
   Legitimacy of Need ............................................................................................... 11
   Physiological Measures and Compliance in Teams ........................................... 13

4. **HYPOTHESES** ................................................................................................. 17

   Hypotheses ........................................................................................................... 17
# Table of Contents (Continued)

## V. METHOD
- Participants ................................................................. 20
- Simulation ........................................................................ 20
- Subsystems ................................................................. 21
- Condition ...................................................................... 22
- Backup Behavior ....................................................... 22
- Physiological Measurement ...................................... 23
- Heart Rate Variability ............................................... 23
- Galvanic Skin Response ............................................. 24
- Team Performance ..................................................... 24

## VI. PROTOCOL
- Pilot Study ................................................................. 25
- Study Protocol ............................................................ 25

## VII. DATA ANALYSIS .................................................. 27

## VIII. RESULTS ............................................................. 31

## IX. DISCUSSION ............................................................ 35
- Limitations ................................................................. 37
- Future Research ........................................................ 40
- Practical Implications ............................................... 41

## X. CONCLUSION .......................................................... 43
Table of Contents (Continued)

TABLES................................................................................................................................. 43

1: Hypotheses and Corresponding Analyses................................................................. 44
2: Correlation Table ..................................................................................................... 46
3: ANOVA of Usage of Monitoring ........................................................................ 47
4: Descriptive Statistics .............................................................................................. 48
5: Physiological Compliance ....................................................................................... 49
Table of Contents (Continued)

APPENDICES ......................................................................................................................... 51

A: Xplant Simulator .............................................................................................................. 52
B: NASA TLX Workload Questionnaire ........................................................................... 53
C: Usage of Monitoring Screen ....................................................................................... 56
D: Empatica E4 Device ....................................................................................................... 57
E: Safety Indexes and Training List .................................................................................. 58
F: Plant Functions Poster ................................................................................................. 60
Table of Contents (Continued)

<table>
<thead>
<tr>
<th>REFERENCES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>61</td>
</tr>
</tbody>
</table>
CHAPTER ONE

Introduction

Organizations are becoming more and more dependent on teams. Often tasks that need to be completed are too large, complex, or complicated for a single person to complete. Therefore, it takes a group of people working together to achieve goals and complete tasks (Mathieu, Maynard, Rapp, & Gilson, 2008). However, when you add others to the task, they have to interact and this can add a host of potential complications that can impact the effectiveness of the team. Nevertheless they are often dependent on each other to meet collective goals and get the job done. If one person fails to complete the necessary tasks related to their role even if all others are fulfilling their goals, the goals of the team might not be met (Crawford & LePine, 2013). Therefore it is necessary to understand and be able to measure the processes that teams go through to ensure team success (Porter, 2005). This is especially important when teams are working together in high stakes contexts; if we do not understand detailed aspects of teams process of effectiveness, then it could lead to team breakdowns and consequently disastrous results (Wilson, Salas, Priest, & Andrews, 2007; Marks & Panzer, 2004). In instances of breakdowns in teams of soldiers, these poor processes can lead to instances of fratricide or accidental friendly fire (Wilson, Salas, Priest, & Andrews, 2007) and in contexts such as process plant operations team process difficulties could lead to chemical disasters (Marks & Panzer, 2004). Although there are a large number of factors that encompass what makes a team effective, two important closely related processes are mutual monitoring and backing up team member’s behaviors. As tasks are becoming larger and
more complex it is becoming increasingly necessary to occasionally share the task workload of others (Hauland, 2008). However, to accomplish this it is important for teams to be monitoring the team’s actions to know when they need to intervene and back up behavior through helping the other team member with their tasks (Salas, Simms, & Burke, 2005). Research has shown that when team members monitor each other and consequently back up behaviors when there is a legitimate need for help, teams have better performance outcomes (Marks & Panzer, 2004; Porter, Hollenbeck, Ilgen, Ellis, West, & Moon, 2003). There can be some complications with these processes however; if a team allocates much of their cognitive resources toward monitoring the team functions it can take away from their effectiveness (Robert & Hockey, 1997). Therefore if there was a way to help automate this process it could be beneficial for the team. Additionally, to this point team monitoring consists of the team being able to observe direct actions (e.g., task performance; Marks & Panzer, 2004). The purpose of this study is to examine if mutual monitoring assistance relates positively to backup behaviors and overall team performance. This could provide implications for the automation of mutual monitoring in teams. To provide a better understanding of what is involved in these processes a review of team effectiveness models and the theoretical background of these constructs is provided.

**Team effectiveness**

The study of teams and their overall effectiveness has been studied through an input process output model (I-P-O) and later input, mediator, output, input (IMOI) model (Ilgen, Hollenbeck, Johnson, & Jundt, 2005). In the I-P-O model Input can be
characterized by things such as individual level factors, (e.g., personality, knowledge, skills, and abilities), group level factors (e.g., size and how the team is organized), and environment level factors (e.g., organizational support and characteristics of the task; Hackman, 1987). Team processes function as a mediator of the relationship between the input and output of a team (Hackman, 1983). According to Marks, Mathieu, and Zacarro (2001) team process occurs in multiple phases that contain transition phase processes, action phase processes, and interpersonal processes (e.g., conflict management).

Transition phases “are periods of time when teams focus primarily on evaluation and/or planning activities to guide their accomplishment of a team goal or objective,” (Marks, Mathieu, & Zaccaro, 2001, p. 364) and involve processes such as goal specification and planning. Action phases “are periods of time when teams are engaged in acts that contribute directly to goal accomplishment” (Marks, Mathieu, & Zaccaro, 2001, p. 366) and involve processes such as mutual monitoring and backup behaviors. Interpersonal phases occur throughout team cycles and include processes such as team conflict. As for the output aspect of team effectiveness, this commonly includes team performance which is solely the overall outcome of the team’s actions and the way the team interacts and works together (Salas, Simms, & Burke, 2005). Marks and colleagues (2001) state that the input process output model can occur dynamically over cycles of performance episodes and include the various phases of processes mentioned earlier. The IMOI model takes this cyclical model further and states that the outcomes of the team become inputs for the next episode of team functioning (Ilgen, Hollenbeck, Johnson, & Jundt, 2005).

Additionally, instead of using the process construct as the mediator function of the model
they expand this to all team mediators such as emergent states (Ilgen, Hollenbeck, Johnson, & Jundt, 2005). The study of mutual monitoring and backup behaviors typically considers them either as both mediators or with mutual monitoring as an input and backup behaviors, as the mediator for team effectiveness (Marks & Panzer, 2004, Marks Mathieu & Zaccaro, 2001). In this study, monitoring assistance condition will take the place of the input, backup behaviors will take the place of the mediator, and performance will function as the outcome.
CHAPTER TWO

Mutual Monitoring and Backup Behaviors

To truly be able to understand backup behaviors and mutual monitoring a review of the theoretical framework is necessary. The various critical team characteristics, emergent states, and processes that play an important role in their foundation will be discussed along with moderators of these behaviors.

Framework

There are a few main processes and team cognitions that characterize backup behaviors and mutual performance monitoring, which are: shared or team mental models, transactive memory systems, and situation awareness (Salas, Simms, & Burke 2005; Salas, Prince, Baker & Sherstha, 1995). If a team has shared mental models and transactive memory systems they are able to be aware of the team interactions, roles, and situation (Mohammed, Ferzandi, & Hamilton, 2010). When a team is aware of the situation they are able to monitor each other and therefore back up other team members when necessary.

Shared mental models and transactive memory systems. First off are the structural aspects of team cognitions that are a key basis for processes such as team monitoring and backup behaviors. These are shared mental models and transactive memory systems (Salas & Cannon-Bowers, 2001). Shared mental models are extraordinarily important and are, team members’ shared, organized understanding of knowledge about key elements of the team’s relevant environment (Klimoski & Mohammed, 1994). These things can be
summarized as the team’s mission and goals and what tasks are required for team members to coordinate and reach these goals in order to achieve the overall mission. In previous research team shared mental models have been differentiated into task mental models and team mental models (Mohammed, Ferzandi, & Hamilton, 2010). Task mental models rely around what specific things are involved in the task, what steps must be taken, and what actions and behaviors are necessary for those steps (Espevik, Johnsen, & Eid, 2011). However, team shared mental models describe the ways the team needs to interact/how interdependent they need to be, which ways they coordinate, and how often they need to coordinate and with whom. Another way that team mental models are characterized is by their similarity and their accuracy (Mohammed, Ferzandi, & Hamilton, 2010). Similarity is how much overlap exists between individual mental models in the team; whereas accuracy is how much overlap exists between the team’s mental model and the mental model(s) of (an) expert(s) (Wildman, Salas, & Scott 2014). Research by Marks et al. (2002) explored how shared team mental models related to backup and found that shared mental models had a positive relationship with backup behavior quality and quantity.

Along with these mental models are transactive memory systems, which include team members being knowledgeable of and able to utilize the cognitive divisions of knowledge, expertise, and skills between members that could be applicable to the job. Essentially, members are aware of each other's mental models and how to work with and share that knowledge (Wegner, 1987; Lewis, 2004). This includes all members of the team being aware of what each team member knows and is capable of, including areas of
expertise and specialization, and are able to call on the correct person when are necessary. Transactive memory systems can be found to be more evident when a team trains together (Liang, Moreland, & Argote, 1995). This also plays into team familiarity; in fact, Smith-Jentsch, Kraiger, Cannon-Bowers, & Salas (2009) found that when teams have experience working together the resulting transactive memory systems lead to increased levels of backup behaviors. Additionally, they found that when team members have higher levels of consensus on what other team members know they had higher levels of backup behaviors.

**Situation awareness** is an emergent state that is related to mutual performance monitoring and backup behaviors. Situation awareness involves environmental perceptions, and the combination of what those perceptions mean relating to the near past, present, and near future (Endsley, 1988; Endsley, 1995). In the team context this involves “teams’ understanding of the situation at one point in time” (Cooke, Kiekel, & Helm, 2001, p. 299). Essentially, teams are dynamically perceiving and understanding similar mental models, preceding actions that got them there, and where the current situation might go in terms of the actions and outcomes of team interactions and task behaviors (Wildman, Salas, & Scott, 2014). In relation to backup behaviors and team monitoring, this is considered to occur after transactive memory systems and shared mental models. To be aware of a situation you have to know information about the team and task contest. This is derived from shared mental models and transactive memory systems (Wildman, Salas, & Scott, 2014). Roth, Multer, and Raslear (2006) found that
situation awareness is a vital part of mutual performance monitoring which leads to backing up behavior.

**Mutual Performance Monitoring**

Mutual performance monitoring essentially involves a team member’s capacity to keep track of fellow team members’ work while also carrying out their own, making sure everything is running as anticipated, and following procedures properly (McIntyre & Salas, 1995). There has not been much research on mutual monitoring and Marks and Panzer (2004) did one of the first studies looking at mutual monitoring in relation to teams. In the past, monitoring has been examined as leadership aspect or an individual self-regulation aspect; however, they brought the concept into the context of teams. A team member being able to know where a team is failing and succeeding is a critical part of mutual performance monitoring. Mutual monitoring is a vital part of team processes and is positively related to processes such as backup behaviors and coordination (Marks & Panzer, 2004). Salas, Simms and Burke (2005) stipulate that backup behaviors are considered the actions taken based on mutual monitoring. Overall mutual monitoring had been found to positively relate to team performance, which is partially due to backup behaviors, and coordination (Marks & Panzer, 2004). Additional aspects of monitoring to consider is that there are only a certain amount of cognitive resources that are available to individuals and when task becomes too complex mutual performance monitoring should be a peripheral action (Roth, Multer, & Raslear, 2006). Additionally, traditional mutual monitoring is only based in observable actions (Marks & Panzer, 2004).
Backup Behaviors

Backup behaviors are essentially the actions that arise from mutual performance monitoring (Salas, Simms, and, Burke, 2005). Research has shown that backup behaviors can include a team member coaching or providing verbal feedback to another member, helping a team member behaviorally by carrying out actions, or completely assuming a task for a teammate (Marks, Mathieu, & Zaccaro, 2001). Although one might think that helping actions for the team could be beneficial whenever they are enacted, studies have shown that that only when there is a legitimate need for backup behavior do these helping behaviors result in increased team effectiveness (Porter et al., 2003; Porter, Gogus, & Chien-Feng Yu, 2010). In fact, it has been found that when behaviors of backup are performed during times of no legitimate need there is actually a reduction in effectiveness (Barnes, Hollenbeck, Wagner, DeRue, Nahrgang, & Schwind, 2008). Thus it is important for the aspects of the legitimacy of need to be considered with backup. In previous research this has been found to appear in disproportionate workload differences. This doesn’t just mean that team members have differences in workload; the disproportionate aspect of workload is the key (Porter, Gous, & Yu, 2010; Porter, Hollenbeck, Illgen, Ellis, West, & Moon, 2003). Porter and colleagues (2003) found when there was a legitimate need and an individual that was receiving backup actions was conscientious or extroverted there were increased backup behaviors over those evident when there was only legitimacy of need. Additionally, when the individual was the provider of backup behaviors the combination of conscientiousness or emotional stability and legitimacy of need predicted higher frequency of backup behaviors beyond only the legitimacy of need.
CHAPTER THREE

Workload Measurement

Mental Workload

Mental workload is becoming an increasingly important construct to study. This is especially true as jobs are switching to more cognitively based tasks (Young & Stanton, 2005). Mental workload can be defined as the amount of attentional resources that is necessary for both objective and subjective performance criteria to be met, of which task demands, external support, and past experience can partially describe the outcome (Young & Stanton, 2001). This draws from the ideas of attentional resource theory where there is only a limited level of cognitive resources available for tasks and when the task exceeds that mental capacity performance decreases unless a change in strategy occurs to reduce this discrepancy (Wellford, 1978; Singleton, 1989). There are multiple methods to measure mental workload. The three main categories are performance based measures, subjective measures, and physiological measures (Meshkati, Hancock, & Rahimi, 1992; Rubio, Diaz, Martin, & Puente, 2004). Performance based measures are founded in the idea that as a task difficulty increases cognitive resource demand will increase and consequently individual performance will decrease. Essentially, performance will be lower in times of high workload and will be higher in times of low workload (Rubio, Diaz, Martin, & Puente, 2004; Young & Stanton, 2005). Subjective workload is based in perceptions of difficulty or effort of a task as a measure of mental workload and typically done through some form of survey given after a performance episode or set of performance episodes (Rubio, Diaz, Martin, & Puente, 2004; Hart & Staveland, 1988).
Lastly are physiological measures of mental workload, which assume that the exertion of mental effort can be measured through various physiological measures (Wilson & Russel, 2003; Kramer, 1990).

**Legitimacy of Need**

For the purposes of using a measure of workload for mutual monitoring to facilitate backup, a mental workload measure that can best associate with legitimacy of need is necessary. As stated earlier when teams perform backup behavior in times of legitimate need they are more effective. Since mutual monitoring is the process that enables backup, if a team is able to have a method of monitoring that captures legitimate need that will have more effective backup behaviors.

Legitimate need is considered by Porter and colleagues (2003) as “when a team member is faced with a higher level of task demands, but has not been compensated with extra resources” (p. 393); this was manipulated through disproportionate workload for one team member as discussed earlier. Barnes and colleagues (2008) went on to look at how a team member monitoring the legitimacy of need of an overloaded team member increased the backup recipient’s dependence on backup. Porter and colleagues (2010) went on further to review the disproportionate workload aspect of legitimacy of need route but included time. They found that time performing in a disproportion of task load influences the legitimacy of need, where legitimacy of need decreases over time. They theorized that this is due to the task work skill increase and the automation of certain tasks. Mental workload as a measure however, should add to more beyond just the task
load as the lack of resources for a task in times of legitimate need are also cognitive (Porter et al., 2010).

**Physiological Measures and Compliance in Teams**

There have been numerous studies that have associated workload, stress and strain with various physiological measures such as electrocardiographic activity, heart rate variability (HRV; calculated through interbeat intervals (IBI)) and Galvanic Skin Responses (GSR; measured through electrodermal activity (EDA)). Additionally, there have been a few studies that have been done on physiological measures of teams which have implications that indicate that physiological measures are appropriate for measurement in teams. For example, measures such as physiological compliance which is “the correlation of physiological measures of team members over time” and team autonomic activity which is “sympathetic and parasympathetic nervous system activity combined” have been found to predict team performance overall (Elkins, Muth, Hoover, Walker, Carpenter, & Switzer, 2009, p. 997). However, since team performance is an outcome which happens in episodes mediated by processes it is important to also physiologically measure the processes of teams along with the outcomes. There’s is also evidence that this is an applicable area too as (Stevens, Galloway, Wang, & Berka 2012) found that electroencephalography or EEG (another physiological measure of workload) complemented measures of team cognition. Of these areas the similarity of individuals’ physiological measurements such as HRV and GSR in the process and performance is particularly apt as they are appropriate single measures over time. Additionally there have been implications that physiological compliance is an indicator of cognitive
readiness (Walker, Muth, Switzer, Rosopa, 2013). This would allow for better cognitive readiness and ability to perceive others’ heavy workload and be able to back up behaviors.

The first of these is a measure of electrocardiographic activity: heart rate variability (HRV). Heart Rate Variability is essentially a measurement over time of the period between consecutive heartbeats or interbeat intervals (Bernston et al., 1997). Another physiological measure associated with workload is Galvanic Skin Response (GSR) which is the electrical conductance of the skin or electrodermal activity (EDA) and can change due to changes in the sympathetic nervous system (Montagu & Coles, 1966). As these have been studied numerous times in the context of workload and stress along with performance their compliance (similarity) could potentially appropriately apply to the team process contexts of mutual monitoring and backup behaviors (Veltman & Gaillard, 1998; Perla & Sterling, 2007; Guhe, Liao, Zhu, Ji, Gray & Schoelles, 2005).

**Heart Rate Variability** In multiple studies, HRV has been found to be linked to large differences between periods of rest (also conceptualized as low workload or stress) and overly heavy mental workload, stress or strain (Veltman & Gaillard, 1998; Hjortskov, Rissen, Blangsted, Fallentin, Lundberg, & Sogaard, 2004). Furthermore, Rowe, Silbert and Irwin (1998) found that HRV indicated when an individual’s ability to complete a task was exceeded. Veltman and Gallard (1998) found that HRV was sensitive to large changes in task difficulty which is potentially a useful differentiation that could examine legitimate need and backup behaviors. The findings of Rowe, Silbert, and Irwin (1998) resulted from looking at teams in five scenarios of varying difficulty and found that HRV
marginally tended to relate to task difficulty; however, when the task went from the second most difficult game to the most difficult game, there was a significant difference in HRV. This indicates that HRV could be a good indicator of cognitive workload overload, again making it an ideal indicator for mutual monitoring of legitimate need for backup. Hjortskov and colleagues (2004) measured HRV through two workload manipulations and found that there was a significant relationship between stressors and HRV. These studies found that HRV was related to psychological stress or workload in a way that is indicative of a workload overload. This exemplifies the type of workload that has been found to associate with contexts where backup behaviors that relate to positive outcomes are theorized to exist and implies that this type of measure could be beneficial if high levels are monitored by the team.

**Galvanic Skin Responses** GSR has also been found to be linked to increased workload and stress (Noubakhsh, Wang, Chen, Calvo, 2012). Noubakhsh and colleagues looked at GSR in relation to two manipulations of workload differences. They found that GSR was significantly related to workload and could additionally differentiate varying levels of workload. Perala and Sterling (2007) measured GSR through three studies and they found that GSR modeled the stress trends that were found in the self-report surveys which provides evidence for its validity. Guhe, Liao, Zhu, Ji, Gray, Schoelles, (2005) looked at GSR through a sensor on a computer mouse in conditions of various task loads where they found a significant difference. A benefit of GSR is that it can clearly be measured over time and is related to task workload as used in backup behavior measurement (Perala & Sterling, 2007). However, it is important to note that there are
numerous ways to capture this data. These measurement techniques have the potential to be utilized for team monitoring and relate to back up and overall team performance. Potential measurement techniques as monitoring methods will now be explored through various hypotheses.
CHAPTER FOUR

Hypotheses

Based in the previously discussed mutual monitoring and backup behavior theory, there are multiple propositions for hypotheses that can be developed based on the method of mutual monitoring in relation to performance and backup behaviors. To explore these hypotheses two conditions were used:

1) The control condition — mutual performance monitoring takes place solely through observation and verbal interactions (note that this is the typical type of informal mutual performance monitoring which often takes place in process control and other teams).

2) The direct monitoring condition — each operator has a monitoring screen which directly presents the performance of the other operators; note that in this condition the typical observation and verbal interaction methods of mutual performance monitoring were available as well. Therefore this condition could also be conceptualized as an "enhanced" mutual performance monitoring condition.

A third condition, direct monitoring of physical state, was not able to be implemented due to technical issues and constraints.

As noted earlier, mutual performance monitoring has been found to be positively related to performance (Marks & Panzer, 2004) which leads to hypothesis one:

H1: Teams that are able to more directly and accurately monitor task performance levels will perform better than teams that can only monitor workload informally.
As there is a positive relationship between legitimate need and backup behaviors (Porter et al., 2008) if teams are able to observe legitimate need through the task performance monitoring assistance window they should be more apt to provide backup:

H2: Teams that are able to directly monitor task performance levels will have more backup behaviors compared to teams that are not able to monitor workload directly.

Since there is a positive relationship between backup behaviors that occur during legitimate need and performance (Porter et al., 2008) in a team with differing workloads backup would theoretically be related to higher performance resulting in hypothesis three:

H3: Teams that have higher numbers of backup behaviors (regardless of MPM type) will have higher performance.

Taking into account the hypothesized positive relationship between the mutual monitoring conditions and performance in times of legitimate need and backup behaviors, along with the positive relationship between backup and performance, a mediate model is hypothesized. Essentially, the expected positive relationship between task performance levels and enhanced MPM is partially explained by the number of backup behaviors:

H4: The positive relationship between mutual monitoring of task workload levels will be partially mediated by backup behaviors.

**Exploratory questions** As discussed earlier, legitimate need generally decreases over time thereby decreasing the necessity of backup behaviors (Porter et al., 2008). Monitoring assistance may exacerbate this effect as teams are able to see the continuous
team member’s performance and adjust accordingly and precisely, which leads to the following exploratory hypotheses:

R1: Performance will increase over time, but with a steeper positive slope in the enhanced MPM condition than the control condition.

Additionally, as the teams are able to see the increase in performance they will be less likely to provide or ask for backup behaviors as they will know if it is necessary or not, which could mitigate the dependence on backup behaviors.

R2: Backup behaviors will be observed earlier in the performance condition than in the control condition.

**Physiological compliance** As discussed above, physiological compliance is likely related to team performance and team cognitions. As the monitoring assistance condition provides information to the team that allows them to see each other’s performance, this could lead to more accurate shared mental models which would, in turn, generate higher physiological compliance. This leads to the final two exploratory research questions:

R3: Physiological compliance among team members will be higher in the mutual monitoring assistance condition.

R4: There will be a positive relationship between physiological compliance and backup behaviors.
CHAPTER FIVE

Method

Participants

Participants were college students recruited from Clemson University. They were recruited through the SONA system for class credit. There were 35 males, 75 females and 1 other. The age of the participants ranged from 18-23, there was a positive skew with a mean of 18.69 and a standard deviation of 1.04. Prior to student arrival the condition of the session (control or experimental) were randomized through a random number generator.

Simulation

This study was conducted using a simulated chemical plant simulator. This simulation is run on computers and meant to resemble a generic chemical plant and very similar to actual operations in chemical plants power plants and other types of process control industries. The two-step goal of the task is to primarily maintain safety of the plant while secondarily producing as much chemical fluid as possible. Chemical fluids run through an overall fluid management system. It begins with an input of fluid entering one subsystem flowing through that system, through the next center subsystem, and lastly through a final subsystem where the fluid is output. It is run by three participants who individually control one of the three subsystems which contain two tanks along with either a heating, cooling, or, catalyst mechanism. The overall model of the organization of the simulation is featured in Appendix A.
**Subsystems**

For each tank in a subsystem, participants are required to manage the fuel, refrigerant, and catalyst in addition to the level and pressure of the fluids in that tank. Fuel, refrigerant, and catalyst are separately managed by the individual participants where one participant manages each fuel/refrigerant/catalyst for their subsystem along with their team member’s subsystems. Tank pressure is adjusted by an individual for their own subsystems through the turning on and off of the vent and pressurizer. Finally the tank fluid level is controlled by the adjustment of the input/output pumps. The way participants know if they need to take action to manage one of the tank contents is by looking at the tanks’ color coded visual cues that indicate the status of the three parameters. If the tank gauges are green that is an indicator that they are at a safe and acceptable level. If the gauges turn yellow that means they need to be attended to and are beginning to get too high or low, and if they turn red that means they are reaching dangerously high or low levels. Additionally, if the center team member would like to relieve control of one of their tanks in their subsystem, another (right or left) member may take over control of that tank on their screen with permission.

**Legitimacy of Need**

To create the disproportionate workload to test the hypotheses we used subsystem B. The participant randomly assigned to operator B experienced very high levels of management difficulty or task workload for three minutes at three occurrences in the simulation. This was intended to result in the one participant having a workload that is
much greater than the other participants creating legitimacy of need. Additionally, to determine legitimate need, when workload begins to increase this was designated as an episode of workload disproportion which was used for the hypotheses. To function as a mental workload legitimacy of need check, the NASA TLX workload measure will be given at the end of the experiment as a check of the workload disproportion (Hart & Staveland, 1988; Appendix B). This is examined through a one way ANOVA between operators on the TLX score.

**Condition**

To manipulate the monitoring condition two differing versions of an indicator screen were located on the bottom left of each participants’ subsystem (Appendix A). In the mutual monitoring assistance condition the screen displayed for each participant the safety performance of the other two participants. In the control condition a non-meaningful power index was displayed for each participant representing the power of the plant for the other participants. To check this manipulation a three-question survey measuring the use of the monitoring screen was given (Appendix C). An example question is “To what degree did you monitor the indicator screen at the bottom left corner during the simulation?” and answer options included a 5-point Likert style scale ranging from “not at all” to “very often”. This scale was found to found to be reliable (α = .80).

**Backup Behavior**

To measure backup behavior, an objective measure was implemented along with subjective behavioral coding. For the objective measure in the experiment, participants were able to take over functions of another team member’s tank if needed. A participant
taking over a tank for another participant was defined as an objective backup behavior for the testing of the hypotheses. Objective backup was quantified through the number of 5-second intervals a tank takeover was in effect. Subjective behavioral coding followed the process used by Marks and Panzer (2004) and used three SME raters trained to distinguish verbal and behavioral backup from other processes such as monitoring and coordination. During training pilot sessions were reviewed and discrepancies resolved. These behaviors included verbal interactions such as “how do I maintain my temperatures” and “can you help me with stabilizing the levels.” Behaviors were quantified by the number of backup behaviors that occurred during the experimental session. Five sessions were also randomly selected and coded by multiple raters to compute interrater reliability and Cronbach’s alpha was used due to the continuous structure and scale of rating (α = .79).

**Physiological Measurement**

Physiological indices were measured through Empatica E4 Wristbands. They are worn around the wrist much like a watch and were unobtrusive (Appendix C). The E4 wristbands measure pulse, HRV, GSR, motion-based activity, and temperature over time. Additionally, there is an indicator button that adds a timestamp that was used to indicate session start and end time for analyzing the data. This over-time measure was used to compute physiological compliance. Physiological compliance was calculated through correlations of physiological members which was then averaged to create a team level compliance score.

**Heart Rate Variability**
HRV is measured at Hz levels which refer to short term changes in blood pressure (Rowe, Sibert, & Irwin, 1998). There are multiple frequency levels; however, the low 0.04-0.15 (typically labeled the 0.1 Hz level) has been used most frequently and deemed applicable for the workload stress or strain context in a social sciences setting (Bernston, et al., 1997; Hjortskov, Rissen, Blangsted, Fallentin, Lundberg, & Sogaard, 2004). When the frequency is lower (closer to 0.15 Hz) that is an indication of low workload or rest whereas when the frequency is higher (closer to 0.4 Hz) that is an indication of very high workload, stress or strain (Rowe, Sibert, & Irwin, 1998).

**Galvanic Skin Response**

GSR can be used as a measure of stress (e.g., from workload) through skin conductance. The way this is examined is through the differences in electrical resistance in the skin where the larger the difference, the higher the stress (Perala & Sterling, 2007). GSR can vary for individuals however, so it is important to note this when comparing GSR between groups (Guhe, Liao, Zhu, Ji, Gray, & Scholles, 2005).

**Team Performance**

Performance of the individuals in the task is measured through how safe the team keeps their tanks. To test the related hypotheses the safety levels are measured every 5 seconds and combined into one overall score through taking the differences at each measurement occasion, squaring them, averaging them, then taking the square root of that to receive an overall score per team member. This score was then averaged to get a team level score, as in the context of this power plant task the outcome of power plant safety is conceptually team based. Due to the process of the calculations of this score, lower
overall scores mean better performance or less safety errors are occurring. With the safety performance an index of pump flow was intended to be used to combine and create overall performance; however, a code complication resulted in the inability to record pump flow. A separate computation with the catalyst output was tested to be used to function as fluid flow performance. To see if using this catalyst based performance measure was appropriate a correlation was conducted with safety. Additionally, safety performance was not only used overall but was calculated the same way but for three separate measurement occasions based around the periods of heavy workload for Operator B. Each safety performance episode included one minute prior to the increased workload, the duration of the increased workload, and three minutes after the increased workload.
CHAPTER SIX

Protocol

Pilot Study

The pilot study of this experiment had teams run through the task with different levels of task workload waves to determine what would result in a task load of legitimate need. Along with this training feedback and adjustments were obtained and adjusted during the pilot study. Additionally, the length of the practice component of training (7 minutes) and length of experimental session (30 minutes) were determined.

Study Protocol

This study began by having the condition pre-selected. Once participants arrived they filled out an informed consent followed by a training that varied on what condition they were in. This included information regarding the task, information based on the monitoring screen with either control (non-significant plant power monitoring screen) or task performance (team performance screen), and information regarding the backup process. They were also provided with a paper displaying the safety indexes (Appendix D) and a poster with the functions of the plant in front of them. Both of these were color coded (Appendix E) Teams then had the opportunity to practice the task for seven minutes to familiarize themselves with the simulation, followed by a thirty minute experimental session. In this session, the task workload increased in three waves for team member B. During the entire session performance, backup and physiological reactions were measured. After the session participants took a short survey consisting of general
demographic questions such as age and sex, the NASA TLX, and usage of monitoring screen.
CHAPTER SEVEN

Data Analysis

The use of correlations, t-tests, and one way and mixed factorial ANOVAs in analyses comparing conditions allowed for determination if the compared groups were significantly different or related for H1-3 and H5-6. Hayes PROCESS is a program that was to be used to test the simple mediation model proposed in H4 if H1-3 were significant. This examines if backup behaviors partially accounted for the relationship between mutual monitoring conditions and performance. Analysis techniques H1-6 are featured in Table 1. As for physiological compliance, three correlations between each team member were computed and then averaged to create a team level variable. These were then compared between conditions and related to backup behaviors as also seen in Table 1 (H7-8).

Of 58 sessions run, only 38 were included in the analyses used for testing the hypotheses. 20 sessions were dropped due to issues such as computer malfunction, colorblindness, and lack of necessary team members. There were 17 control sessions and 21 experimental sessions included. Two measures of task performance were implemented, a safety measure (the primary measure) and a productivity measure (a secondary measure). The safety measure was a composite (at the operator level) of the root-mean-squared deviations of the operator's unit components from the target "safe" levels. Deviations from these target values indicated decreased safety and therefore decreased task performance (as the participants had been informed that safety was their primary goal). These deviations were aggregated across the six unit components (per
operator) and across the relevant time periods. Due to a programming error, the original direct measure of productivity (flow of the liquid product through the entire system) was not available. Instead a proxy measure of productivity, total amount of catalyst used, was used as a substitute.

A direct count and time measure of the "back-up button" implemented in the simulation was used as the primary measure of backup behaviors. However, this variable was seldom used (12 cases) and had a very large overall range (5-1,430 seconds) which made it problematic. As a result the observer-made subject backup behavior measure was used instead. The one situation in which the objective backup behavior variable was used as the primary testing variable was in H6 where it was necessary to have an overtime backup variable which is recorded in the objective measure. We also correlated the backup behavioral measures and they were significantly positively related which gives indication that this variable is still usable for H6 (Table 2). Additionally, since the variable of production was not able to be calculated in the form of pump flow an attempted production variable based on the catalyst was examined by correlating it with safety performance and backup behaviors. This was found to be significantly negative for the subjective backup and safety performance and since this is not a true measure of actual production this variable is not used throughout and only safety performance was used (Table 2).
To check the mutual performance monitoring manipulation a three-question measure of monitoring screen usage was used. The responses were averaged to create an overall monitoring screen usage score. Using all three questions separately along with the overall measure, four one-way ANOVAs were conducted comparing usage of the monitoring screen between the conditions. There was no significant difference in the overall usage measure but for two of the three questions however there was a significant difference in the question “To what degree was the indicator screen at the bottom left corner important for your teams’ performance?” between groups, where the indicator screen was considered significantly more important for team performance in the monitoring condition $F(1, 109) = 4.77, p = .031, \eta^2_p = .042$; see Table 3.

There were a few physiological compliance measurement issues of the HRV and GSR variables and their effect on the analyses run needs to be addressed. First is the variable of HRV, the way the Empatica E4 wristband measured HRV resulted in multiple time gaps due to sensor and/or measurement error as the device does not analyze HRV over sampling periods where there is some kind of error. To compute physiological compliance the correlation of the participants’ physiological measures over the session is required. The issue is however that the calculation gaps vary by person resulting in an inability to have a correlation of the same physiological sampling instances over the session. Consequently, HRV was unable to be used in calculating physiological compliance and the associated hypotheses. This is also appropriate however since Rowe, Silbert, and Irwin (1998) and Veltman and Gallard (1998) found that HRV only perceives large workload differences and the NASA TLX found that at least between individuals
there wasn’t a significant difference in perceived workload. Therefore, this might have not been an appropriate measure in this context.

In relation to GSR, this physiological indicator is measured with the E4 Empatica wristband over consistent time intervals; therefore, it is a more apt variable for computing physiological compliance. One stipulation to using this measure is that any measurement error is included in the analysis as well. However since there are 7,200 measurement occasions during each 30 minute session there should be enough useful information to calculate a physiological compliance measure.
CHAPTER EIGHT

Results

To test H1 and H2, t-tests were conducted between condition and the following variables: overall safety performance, overall subjective backup behaviors, and overall objective backup behaviors. Tests indicated that there was no significant difference in overall safety performance between conditions $t(36) = -1.33, p = .19$ rejecting H1. Additionally, there was no significant difference in overall subjective backup behaviors between conditions $t(36) = 1.65, p = .11$ rejecting H2. Along with this, there was no significant difference in overall objective backup behavior between conditions $t(36) = -0.13, p = .90$ means and standard deviations can also be seen in Table 4.

To test H3, a correlation was computed between subjective backup behaviors and overall safety performance and found to be non-significant, rejecting H3 (Table 2). Additionally, there was no significant correlation between overall objective backup behaviors and overall safety performance (Table 2).

Since no significant relationships between condition, backup behaviors, and performance were found in the first three hypotheses the mediation in H4 is rejected due to no relationship to mediate.

For R1 a 3X2 mixed factorial ANOVA was calculated to examine if there is increased safety performance over time where that relationship is stronger in the performance monitoring condition. A main effect of time was nonsignificant $F (2, 72) = .71, p = .50, \eta_p^2 = .019$ along with a main effect of condition being non-significant $F (1,
36) = 2.42, \( p = .13, \eta_p^2 = .063 \) therefore the interaction was also insignificant \( F (2, 72) = 1.43, p = .25, \eta_p^2 = .038 \), rejecting H5.

For R2 a 3X2 mixed factorial ANOVA was also calculated. This examined if objective backup behaviors would be higher in earlier time periods in the performance monitoring condition. A main effect of time was nonsignificant \( F (2, 72) = .16, p = .90, \eta_p^2 = .004 \) along with a main effect of condition being insignificant \( F (1, 36) = .09, p = .77, \eta_p^2 = .002 \) therefore the interaction was also insignificant \( F (2, 72) = 1.73, p = .19, \eta_p^2 = .046 \), rejecting R2.

For calculating physiological compliance correlations were run between measurement occasions of GSR for each pair of operators in a team. These correlations were then averaged to make a team level index of physiological compliance. Individual pairs and team level physiological compliance scores long with group averaged totals and operator average totals are included in Table 5. To examine if physiological compliance is higher in the monitoring assistance condition for R3 an independent samples t-test was calculated to compare physiological compliance across conditions. There was no significant difference in physiological compliance between conditions resulting in the rejection of R3 \( t(34) = -.83, p = .42 \). To examine the relationship between backup behaviors and physiological compliance the team level physiological compliance score was correlated with subjective backup behavior scores resulting in no significant relationship \( r = -0.86, p = .39 \) rejecting R4.

To check the manipulation of monitoring condition a three-question measure of monitoring screen usage was used. The responses were averaged to create an overall
monitoring screen usage score. Using all three questions separately along with the overall measure, four one-way ANOVAs were conducted comparing usage of the monitoring screen between the conditions. There was no significant difference in the overall usage measure and two of the three questions; however, there was a significant difference in the question “To what degree was the indicator screen at the bottom left corner important for your teams’ performance?” between groups, where the indicator screen was considered significantly more important for team performance in the monitoring condition $F(1, 109) = 4.77, p = .031, \eta^2_p =.042$; see Table 3.
CHAPTER NINE

Discussion

The hypothesized results of this study were found to be non-significant; however there are some interesting implications and questions that can be derived from the research. When looking at the usage of monitoring screen questions, the one question relating to importance of the screen for team performance was significant. This has potential implications for how participants perceive the presence of a performance monitoring screen. Participants’ perception of a mutual monitoring screen’s importance for their team performance could result in some interesting relationships with team processes and emergent states. For example, this could be related to improved shared mental models. The perceived performance importance of monitoring each other could relate to a higher emergent shared mental model of their team performance in a task. The other two questions however were non-significant. One indication from this non-significant result is that teams simply did not contribute a significant amount of cognitive resources towards the monitoring of each other’s performance in comparison to a non-relevant indicator. In the context of mutual monitoring though this could be a positive thing as according to Roth and colleagues (2006) monitoring should be a passive activity that is not a primary task. This potential effect should be further explored.

When looking at hypotheses one and two, experimental condition did not have a significant effect on backup or performance. This does not support the hypotheses that the monitoring assistance condition positively affects team performance and backup behaviors. This could implicate that teams are effectively able to monitor each other
without assistance. By being face to face, teams may be able to provide all of the mutual monitoring necessary for team performance and backup by interacting and communicating with each other.

For hypothesis three, no positive relationship was found between backup behaviors and performance. As the manipulation of workload to create legitimacy of need was non-significant, this non-significant relationship between backup behaviors and performance could provide support for the findings of Porter and colleagues (2008) that backup behaviors are only beneficial in legitimate times of need. This could provide evidence for a focus on creating a team understanding of what constitutes a legitimate and non-legitimate need when coming to the aid of another teammate. Due to the non-significance of an effect of condition on performance or backup behaviors and backup behaviors on performance hypothesis 4 was not calculated. With no relationships between these variables there is no testable relationship between mutual monitoring of performance that can be partially explained by backup behaviors.

When looking at research question 1, there was no significant main effect of timing episodes when looking at performance along with no significant effect of condition. Consequently, there was no interaction between condition and timing episodes on performance. This could have implications for how much time it takes for a team to have performance improvement in specific contexts. As this was a 30-minute task with a relatively complex task this could support evidence that Marks and Panzer (2004) provided that the automaticity of tasks that occurs over time resulting in better performance takes longer for more complex tasks.
Research question 2 was also found to be non-significant as timing episodes or condition did not have an effect on backup behaviors. Consequently, there was no interaction between condition and timing episodes on backup behaviors. This non-significance can potentially have implications about the characterization of legitimacy of need where backup behaviors would be appropriate. Maybe simply a task’s novelty isn’t necessarily a large component of legitimacy of need and therefore task workload would potentially be prioritized when studying this phenomenon.

Research questions three and four were also found to be non-significant. This is interesting as the differences in operator pair physiological compliance could potentially be indicative of intra-group differences. If one group has one pair with a strong positive correlation between two members and a weak correlation of physiological compliance between those two members and the other team member there could be some team characteristics that resulted in this intra-group difference. It is interesting to see this as physiological compliance has traditionally been studied in two person teams and this adds a depth to our understanding to the possibilities of using physiological compliance in our study of people in teams.

Limitations

There are several limitations for this study that could have greatly influenced the results. These include the lack of a significant workload manipulation, the computing issue with the performance variable, potential task complexity issues and lack of training comprehension, the issue with measurement and use of objective vs subjective backup
behaviors, the calculation issue of HRV for physiological compliance, and measurement issues of GSR.

**Workload** The nonsignificant difference between operators in this study resulted in calculation limitations and potentially had a large effect on the results of this study. As legitimacy of need is such an important part of backup behavior that benefits team performance, not having this legitimate need could have been a large factor in the non-significance of our results. With no manipulated legitimate need that teams could monitor, teams would not be as likely to use backup when it was not necessary. Also this reduced the theoretical ability to compare backup behavior quality based in team members’ differential experience of times of legitimate vs non-legitimate need.

**Performance** The variable of performance had a limiting factor in that it wasn’t a complete measure of all facets of the team performance. The safety of the team was captured; however the productivity was unable to be captured due to a computing error. As the goals of the study were (in order) safety, followed by efficiency, the lack of that aspect of performance creates a deficiency of that construct. This could possibly have affected the results of the study, for if the productivity was considered in combination with safety, more of the variance in back up could have potentially been explained by the monitoring condition or backup behaviors.

**Task complexity and Training** After reviewing the NASA TLX and through the behavior coding this task was perceived as difficult overall for participants. Due to the nature of the task, if there is a complete understanding of the functions of the plant this task is less difficult. However if there is not a complete understanding of the task the
complexity of the numerous different facets that need to be controlled can create a higher level of difficulty. Even though there was a complete PowerPoint training and the opportunity to practice with objectives intended to familiarize them with the simulation there was a lot of information to be observed at once. There was not opportunity to intermittently learn about each aspect of the simulation and practice with that information immediately afterward before moving on to the next thing. This could have partially contributed to the finding of no significant difference of perceived workload between operators.

**Objective backup behaviors** Another limitation is related to backup behaviors. The objective backup behavior was seldom used and when it was, the usage varied greatly; therefore, it was unable to be used in combination with the subjective backup behaviors. This resulted in only the subjective backup measure used for the majority of hypothesis testing. The way that this method for backing up behavior was structured (fully taking over tanks) could have felt invasive and led to a reduction in the desire to use it over a subjective interactional method of backup.

**HRV** HRV in this context was unable to be used in computing physiological compliance due to the measurement sampling method and insignificant difference of workload between operators. As HRV has been theoretically linked to cognitions and team performance, lacking this variable in calculations results in an inability to have a more comprehensive picture of physiological compliance.

**GSR** When measuring GSR there could have been a number of measurement errors. This may have resulted in incorrect measurement of actual skin responses of individuals which
could have potentially weakened the correlations and consequential physiological compliance that was calculated. This potentially could have contributed to the non-significance found when testing R3 and R4.

**Future research**

Using a simulation for a chemical plant function is a very realistic scenario as industrial tasks become more computer based. However, what if that is taken a step further? Our team was face to face. However, providing team monitoring assistance should also be studied in a virtual team setting. As we found that the level of usage of the performance monitoring screen was not significantly different between conditions, the participants were able to be face to face and monitor each other through verbal and visual cues. In a virtual team, the verbal and visual cues that would indicate that a teammate is in trouble would be less apparent, so therefore a performance monitoring screen could be much more valuable and useful in this context.

In this study we only looked at how monitoring assistance through a task workload or performance indicator would contribute to team outcomes and processes. There are however, other ways to measure mental workload such as physiological reactions. HRV and GSR have both been shown to be related to mental cognitive workload. Since often observable actions do not fully represent cognitive processes (Marks & Panzer, 2004) mental workload issues are not as easily monitored by other team members as they are also founded in cognitions not simply in task performance levels (Young & Stanton, 2005). Therefore, monitoring assistance provided through a physiological indicator of mental workload could potentially create differing results.
When helping in a team, one emergent state that could be researched in the context of mutual monitoring and backup behavior is trust. To accept the help of another team member should mean that that team member trusts the other team member to have their best interests in mind and know how to help. Additionally, looking at trust related to mutual monitoring, a team who doesn’t trust each other could potentially place a higher cognitive load on monitoring their other team members. Also the providing of backup could be related to a team member not trusting the member that is receiving backup to be able to complete the task. There are a multitude of contexts where trust should be theoretically examined in relation to these team processes.

When looking at the physiological compliance between teams there is great potential for using compliance between pairs of operators to examine intra-team differences. This could be potentially examined in context to team member differenced in processes or emergent states. It should be looked at to examine if physiological compliance pair differences indicate differences in mental models, trust or a great number of other team variables. As this is a relatively new area of research there are many potential directions that this can be researched.

**Practical implications**

Findings from this study could inform teams in practice about the importance of legitimacy of need. Teams with this knowledge could provide training about knowing when to back up behavior. Information should be provided detailing how, when backing up behavior, it is important that the person receiving backup legitimately need it while the person providing the backup also has a lower workload as to not take away from their
task. This could put performance at a standstill or even decrease performance if teams do not implement this process to backing up behavior.

For face-to-face teams understanding that monitoring through interactions could be a very effective way of ensuring that the process that is occurring has benefits for team norms and policy. If it is understood in a team that checking in by saying something such as “how are your levels” might be better than looking at the levels; this can result in less confusion and conflict. When a question is asked, an individual’s perception of how they are managing the levels can be included in the response where just looking at a level doesn’t have the context, just the content related. Therefore if a face-to-face team forms norms and policy around verbal monitoring this can prove beneficial.
CHAPTER TEN

Conclusion

Overall, none of the proposed hypotheses were supported; however, this study did contribute some interesting information about face-to-face teams’ mutual monitoring, the perceived relevancy of a monitoring screen, and the legitimacy of need for backup behaviors. This study was in a face-to-face team where members could interact and receive verbal and visual cues. This study has implications that in a team where people are directly face to face the interactional mutual monitoring might be sufficient. However, many teams are not constantly face to face and this effect could be different in those contexts. There was one significant difference which was the perceived relevancy of the performance monitoring screen for team performance. These perceptions could relate to additional team emergent states and processes as their perception could result in the team taking additional actions that they wouldn’t otherwise. Additionally, as this study did not have a difference in workload between operators, this resulted in a lack of legitimacy of need. Since backup behaviors have been theoretically found to be only beneficial in a time of legitimate need the lack of significance between backup behaviors and performance supported this effect. In relation to physiological compliance, the insignificant difference could factor into some intra-team differences between the correlations between pairs of operators. This theoretical framework has multiple opportunities to be researched further in differing contexts, related to additional team emergent states and processes, in teams with differences in legitimate need, and with varying monitoring assistance workload indices.
Table 1

Hypotheses and Corresponding Analyses

Table 1.

<table>
<thead>
<tr>
<th>Study Variables, Hypotheses and Corresponding Analyses</th>
<th>Variables</th>
<th>Hypothesis</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IV Condition</td>
<td>H1 Teams that are able to monitor task performance levels will perform better than teams that are not able to monitor workload</td>
<td>T test to compare means between groups</td>
</tr>
<tr>
<td></td>
<td>DV Performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IV Condition</td>
<td>H2 Teams that are able to monitor task performance levels will have more backup behaviors compared to teams that are not able to monitor workload</td>
<td>T test to compare means between groups</td>
</tr>
<tr>
<td></td>
<td>DV Backup behaviors</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IV Backup behaviors</td>
<td>H3 Teams that have more backup behaviors will have higher performance</td>
<td>Correlation between performance and backup behaviors</td>
</tr>
<tr>
<td></td>
<td>DV Performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mutual monitoring, backup behaviors, and control</td>
<td>H4 The positive relationship between mutual monitoring of task workload levels during legitimate need will be partially explained by backup behaviors</td>
<td>Simple mediation using Hayes PROCESS</td>
</tr>
<tr>
<td></td>
<td>IV Condition</td>
<td>R1 Performance will increase over time with a steeper slope in</td>
<td>3X2 mixed factorial ANOVA</td>
</tr>
<tr>
<td></td>
<td>IV Timing episodes</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DV Performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV Condition</td>
<td>IV Timing episodes</td>
<td>DV Backup behaviors</td>
<td>Analysis Method</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------------</td>
<td>---------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>R2 Backup behavior</td>
<td>Backup behavior</td>
<td>3X2 mixed factorial ANOVA</td>
<td></td>
</tr>
<tr>
<td>will be higher in</td>
<td>earlier times in the</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>performance condition</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>and will be steadily low in the control condition</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IV Condition</th>
<th>DV Physiological compliance</th>
<th>Analysis Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>R3: Physiological compliance will be higher in the mutual monitoring assistance condition</td>
<td>T test to compare means between groups</td>
<td></td>
</tr>
</tbody>
</table>

| Physiological compliance and backup behaviors | R4 There will be a positive relationship between physiological compliance and backup behaviors | Correlation between physiological compliance and backup behaviors |

*Note. Table includes all hypothesized relationships and analysis methods*
Table 2

Correlation table

Table 2.

Correlations between Variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Overall Safety Performance</th>
<th>Production Performance</th>
<th>Subjective Backup Behaviors</th>
<th>Objective Backup Behaviors Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Safety Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Production Performance</td>
<td>-.368*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjective Backup Behaviors</td>
<td>-.146</td>
<td>.242</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objective Backup Behaviors Overall</td>
<td>.308</td>
<td>.047</td>
<td>.333*</td>
<td></td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.05 level (2-tailed).
Table 3

ANOVA of Usage of Monitoring

Table 3.

ANOVA Condition on Usage of Monitoring Scale

<table>
<thead>
<tr>
<th>Question</th>
<th>df</th>
<th>F</th>
<th>$\eta^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>To what degree did you monitor the indicator screen at the bottom left corner during the simulation?</td>
<td>1</td>
<td>.607</td>
<td>.628</td>
<td>.438</td>
</tr>
<tr>
<td>When carrying out the simulation, what level of importance did you place on the indicator screen at the bottom left corner?</td>
<td>1</td>
<td>1.081</td>
<td>1.216</td>
<td>.301</td>
</tr>
<tr>
<td>To what degree was the indicator screen at the bottom left corner important for your teams’ performance?</td>
<td>1</td>
<td>4.768</td>
<td>7.560</td>
<td>.031</td>
</tr>
<tr>
<td>Overall usage of monitoring screen</td>
<td>1</td>
<td>2.684</td>
<td>2.396</td>
<td>.104</td>
</tr>
</tbody>
</table>

*Note. This table includes the ANOVA of condition on usage of monitoring screen.*
Table 4

Descriptive Statistics

<table>
<thead>
<tr>
<th>Descriptors</th>
<th>Mean Control</th>
<th>Std. Deviation Control</th>
<th>Mean Experimental</th>
<th>Std. Deviation Experimental</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 1 Safety Performance</td>
<td>0.29</td>
<td>0.17</td>
<td>0.24</td>
<td>0.11</td>
</tr>
<tr>
<td>Time 2 Safety Performance</td>
<td>0.29</td>
<td>0.25</td>
<td>0.18</td>
<td>0.11</td>
</tr>
<tr>
<td>Time 3 Safety Performance</td>
<td>0.31</td>
<td>0.31</td>
<td>0.18</td>
<td>0.17</td>
</tr>
<tr>
<td>Subjective Backup Behaviors</td>
<td>12.71</td>
<td>11.22</td>
<td>21.95</td>
<td>20.779</td>
</tr>
<tr>
<td>Objective Backup Behaviors Overall</td>
<td>18.82</td>
<td>69.31</td>
<td>16.71</td>
<td>31.60</td>
</tr>
<tr>
<td>Time 1 Objective Backup Behaviors</td>
<td>3.35</td>
<td>13.82</td>
<td>6.47</td>
<td>17.72</td>
</tr>
<tr>
<td>Time 2 Objective Backup Behaviors</td>
<td>6.06</td>
<td>20.80</td>
<td>1.81</td>
<td>7.84</td>
</tr>
<tr>
<td>Time 3 Objective Backup Behaviors</td>
<td>5.18</td>
<td>17.85</td>
<td>2.57</td>
<td>8.83</td>
</tr>
</tbody>
</table>

Note. This table includes means and standard deviations for the variables used in H1-6 and R1 & R2.
Table 5

Physiological Compliance Table

Table 5.

Physiological Compliance

<table>
<thead>
<tr>
<th>Team</th>
<th>Condition</th>
<th>AB</th>
<th>BC</th>
<th>AC</th>
<th>PhysioCompl</th>
<th>Averages of Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>808</td>
<td>1</td>
<td>-0.58864</td>
<td>-0.36272</td>
<td>0.576809</td>
<td>-0.12485133</td>
<td></td>
</tr>
<tr>
<td>804</td>
<td>1</td>
<td>-0.21265</td>
<td>-0.21265</td>
<td>-0.14784</td>
<td>-0.19104667</td>
<td></td>
</tr>
<tr>
<td>802</td>
<td>1</td>
<td>0.845958</td>
<td>0.371526</td>
<td>-0.08287</td>
<td>0.37820433</td>
<td></td>
</tr>
<tr>
<td>800</td>
<td>1</td>
<td>0.072262</td>
<td>0.083417</td>
<td>0.167515</td>
<td>0.10773133</td>
<td></td>
</tr>
<tr>
<td>790</td>
<td>1</td>
<td>0.083454</td>
<td>0.327816</td>
<td>-0.34011</td>
<td>0.02372</td>
<td></td>
</tr>
<tr>
<td>784</td>
<td>1</td>
<td>0.032335</td>
<td>0.178015</td>
<td>0.053183</td>
<td>0.08784433</td>
<td></td>
</tr>
<tr>
<td>782</td>
<td>1</td>
<td>0.088736</td>
<td>-0.12608</td>
<td>0.14784</td>
<td>0.02108667</td>
<td></td>
</tr>
<tr>
<td>771</td>
<td>1</td>
<td>-0.1689</td>
<td>0.582835</td>
<td>-0.0112</td>
<td>0.13424467</td>
<td></td>
</tr>
<tr>
<td>758</td>
<td>1</td>
<td>0.173595</td>
<td>0.032872</td>
<td>-0.60786</td>
<td>-0.13379633</td>
<td></td>
</tr>
<tr>
<td>754</td>
<td>1</td>
<td>0.305733</td>
<td>-0.02273</td>
<td>-0.09509</td>
<td>0.06263833</td>
<td></td>
</tr>
<tr>
<td>712</td>
<td>1</td>
<td>0.023481</td>
<td>-0.21587</td>
<td>-0.08741</td>
<td>-0.093265</td>
<td></td>
</tr>
<tr>
<td>703</td>
<td>1</td>
<td>0.036241</td>
<td>0.212803</td>
<td>0.161752</td>
<td>0.136932</td>
<td></td>
</tr>
<tr>
<td>700</td>
<td>1</td>
<td>0.173486</td>
<td>-0.14893</td>
<td>-0.19692</td>
<td>-0.05745467</td>
<td></td>
</tr>
<tr>
<td>698</td>
<td>1</td>
<td>0.010116</td>
<td>0.004521</td>
<td>0.282895</td>
<td>0.09917733</td>
<td></td>
</tr>
<tr>
<td>692</td>
<td>1</td>
<td>-0.15341</td>
<td>-0.49917</td>
<td>0.465418</td>
<td>-0.06238733</td>
<td></td>
</tr>
<tr>
<td>688</td>
<td>1</td>
<td>-0.04542</td>
<td>-0.23979</td>
<td>0.078781</td>
<td>-0.068812</td>
<td></td>
</tr>
<tr>
<td>686</td>
<td>1</td>
<td>-0.11988</td>
<td>0.44228</td>
<td>-0.4682</td>
<td>-0.04859933</td>
<td></td>
</tr>
<tr>
<td>684</td>
<td>1</td>
<td>-0.25399</td>
<td>0.169751</td>
<td>0.232818</td>
<td>0.04952567</td>
<td></td>
</tr>
<tr>
<td>682</td>
<td>1</td>
<td>-0.26628</td>
<td>-0.01009</td>
<td>-0.10871</td>
<td>-0.12836233</td>
<td></td>
</tr>
<tr>
<td>806</td>
<td>0</td>
<td>-0.50261</td>
<td>0.543246</td>
<td>-0.87382</td>
<td>-0.27773</td>
<td></td>
</tr>
<tr>
<td>780</td>
<td>0</td>
<td>0.101085</td>
<td>0.030665</td>
<td>-0.55496</td>
<td>-0.14106833</td>
<td></td>
</tr>
<tr>
<td>778</td>
<td>0</td>
<td>-0.37216</td>
<td>-0.48748</td>
<td>0.144998</td>
<td>-0.238213</td>
<td></td>
</tr>
<tr>
<td>769</td>
<td>0</td>
<td>-0.4602</td>
<td>-0.58875</td>
<td>0.36029</td>
<td>-0.22955067</td>
<td></td>
</tr>
<tr>
<td>760</td>
<td>0</td>
<td>-0.20547</td>
<td>-0.03428</td>
<td>0.186462</td>
<td>-0.017762</td>
<td></td>
</tr>
<tr>
<td>756</td>
<td>0</td>
<td>0.293677</td>
<td>0.072071</td>
<td>0.737932</td>
<td>0.36789333</td>
<td></td>
</tr>
<tr>
<td>752</td>
<td>0</td>
<td>0.058292</td>
<td>-0.34647</td>
<td>0.098815</td>
<td>-0.06312</td>
<td></td>
</tr>
<tr>
<td>747</td>
<td>0</td>
<td>0.043657</td>
<td>0.758026</td>
<td>0.040349</td>
<td>0.28067733</td>
<td></td>
</tr>
<tr>
<td>745</td>
<td>0</td>
<td>-0.55588</td>
<td>0.20179</td>
<td>-0.71198</td>
<td>-0.35535667</td>
<td></td>
</tr>
<tr>
<td>743</td>
<td>0</td>
<td>0.547684</td>
<td>0.68482</td>
<td>0.750623</td>
<td>0.66104233</td>
<td></td>
</tr>
<tr>
<td>738</td>
<td>0</td>
<td>0.375166</td>
<td>0.302163</td>
<td>0.135118</td>
<td>0.27081567</td>
<td></td>
</tr>
<tr>
<td>724</td>
<td>0</td>
<td>-0.36094</td>
<td>-0.36936</td>
<td>0.989882</td>
<td>0.086528</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>-0.45221</td>
<td>-0.4827</td>
<td>0.882284</td>
<td>-0.017543</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>----------</td>
<td>---------</td>
<td>----------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>721</td>
<td>0</td>
<td>-0.4412</td>
<td>-0.68345</td>
<td>0.19877</td>
<td>-0.30862467</td>
<td></td>
</tr>
<tr>
<td>719</td>
<td>0</td>
<td>0.507803</td>
<td>-0.15845</td>
<td>-0.28329</td>
<td>0.022019</td>
<td></td>
</tr>
<tr>
<td>716</td>
<td>0</td>
<td>0.310631</td>
<td>0.795024</td>
<td>0.324593</td>
<td>0.47674933</td>
<td></td>
</tr>
<tr>
<td>710</td>
<td>0</td>
<td>0.036079</td>
<td>0.079775</td>
<td>-0.42142</td>
<td>-0.10185367</td>
<td></td>
</tr>
<tr>
<td>708</td>
<td>0</td>
<td>0.024568</td>
<td>0.05495</td>
<td></td>
<td>0.024406</td>
<td></td>
</tr>
</tbody>
</table>

**Average Operator**  
-0.0289  0.024568  0.05495

**Total Average**  
0.016873

| Backup behaviors correlation | $r = -0.15$ | $p = 0.39$
|------------------------------|-------------|-------------|

**T-tests between conditions**  
$t = -0.83$  $df = 34$  $p = 0.42$

*Note. This table includes all physiological compliance scores for pairs, teams, groups, average operator pair, total average, correlation between backup behaviors and team physiological compliance, and the Fischers Z test for between group physiological compliance.*
Appendix A

Xplant simulator

Figure 1. Visual representation of Xplant Chemical simulator.
Appendix B

NASA TLX Workload Questionnaire

NASA TLX Workload Questionnaire

Workload Survey

Here we are interested in examining the experiences that you think that you will have during the mission. In the most general sense, we are examining the sense of “workload” experienced during the mission(s).

Workload is a difficult concept to define precisely. The factors that influence your experience of workload may come from several factors. This survey is divided into four sections which will serve to assess workload. As two sections deal with assessing perceptions of your workload and two sections deal with assessing your perception of workload, please read the instructions for each section carefully before completing.
**Instructions:** Place an X on each scale at the point that best represents your experience of workload during the mission. Marks must be placed inside the box, not on the lines.

1. **Mental Demand:**
How much mental and perceptual activity did the mission require of you (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)?

<table>
<thead>
<tr>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
</table>

2. **Physical Demand:**
How much physical activity did the mission require of you (e.g., pushing, pulling, turning, controlling, activating, etc.)? This refers to you not your soldier.

<table>
<thead>
<tr>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
</table>

3. **Temporal Demand:**
How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred?

<table>
<thead>
<tr>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
</table>

4. **Performance:**
How successful do you think you were in accomplishing the goals of the task? How satisfied were you with your performance in accomplishing these goals?

<table>
<thead>
<tr>
<th>Bad</th>
<th>Average</th>
<th>Good</th>
</tr>
</thead>
</table>

5. **Effort:**
How hard did you have to work (mentally and physically) to accomplish your level of performance?

<table>
<thead>
<tr>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
</table>

6. **Frustration:**
How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

   | Low | Medium | High |
**Instructions:** For each of the pairs (for example, mental demand vs. effort) choose which one of the two items was more important to your experience of workload. (Circle).

**KEY**

**Effort:** Mental and physical work required to accomplish your level of performance.

**Temporal:** Pressure due to the rate or pace at which the task or parts of the task occurred.

**Physical:** Physical activity required (e.g., pushing, pulling, turning, controlling, activating, etc.).

**Performance:** Satisfaction with your performance.

**Frustration:** Frustration (i.e., insecure, discouraged, irritated, stressed, and annoyed) felt during the task.

**Mental:** Mental and perceptual activity required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.).

<table>
<thead>
<tr>
<th>Effort or Performance</th>
<th>Temporal Demand or Frustration</th>
<th>Performance or Frustration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal Demand</td>
<td>Physical Demand or Frustration</td>
<td>Physical Demand or Temporal Demand</td>
</tr>
<tr>
<td>or Effort</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td>Temporal Demand</td>
<td>Performance</td>
</tr>
<tr>
<td>or</td>
<td>Physical Demand</td>
<td>or Frustration</td>
</tr>
<tr>
<td>Effort</td>
<td></td>
<td>or Temporal Demand</td>
</tr>
<tr>
<td></td>
<td>Physical Demand</td>
<td></td>
</tr>
<tr>
<td>or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td>Temporal Demand</td>
<td>Mental Demand</td>
</tr>
<tr>
<td>or</td>
<td>or</td>
<td></td>
</tr>
<tr>
<td>Mental Demand</td>
<td>Performance or Frustration</td>
<td>or Effort</td>
</tr>
<tr>
<td>or</td>
<td></td>
<td>or Temporal Demand</td>
</tr>
<tr>
<td>Performance</td>
<td>or</td>
<td>or Effort</td>
</tr>
<tr>
<td>or</td>
<td>Mental Demand</td>
<td></td>
</tr>
<tr>
<td>Physical Demand</td>
<td>Effort</td>
<td>Frustration</td>
</tr>
<tr>
<td>or</td>
<td>or</td>
<td>or Mental Demand</td>
</tr>
<tr>
<td>Physical Demand</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figure 2. NASA TLX Workload Questionnaire.*
Appendix C

Usage of Monitoring Screen

Usage of Monitoring Screen

1) To what degree did you monitor the indicator screen at the bottom left corner during the simulation?

<table>
<thead>
<tr>
<th>Response</th>
<th>1 = not at all</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5 = Very often</th>
</tr>
</thead>
</table>

2) When carrying out the simulation, what level of importance did you place on the indicator screen at the bottom left corner?

<table>
<thead>
<tr>
<th>Response</th>
<th>1 = not important</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5 = very important</th>
</tr>
</thead>
</table>

3) To what degree was the indicator screen at the bottom left corner important for your teams’ performance?

<table>
<thead>
<tr>
<th>Response</th>
<th>1 = not important</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5 = very important</th>
</tr>
</thead>
</table>

Figure 3. Usage of monitoring screen survey
Appendix D

Empatica E4 Device

Figure 4. Empatica E4 wrist device
### Appendix E

#### Safety Indexes

<table>
<thead>
<tr>
<th></th>
<th>Great (Ideal)</th>
<th>Good (green)</th>
<th>Okay (yellow)</th>
<th>Bad (red)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Temperature</strong></td>
<td>40</td>
<td>Between 30 &amp; 50</td>
<td>Between 20 &amp; 60</td>
<td>Lower than 20 &amp; greater than 60</td>
</tr>
<tr>
<td><strong>Pressure</strong></td>
<td>0</td>
<td>Between -5 &amp; 5</td>
<td>Between -10 &amp; 10</td>
<td>Lower than -10 &amp; higher than 10</td>
</tr>
<tr>
<td><strong>Tank Level</strong></td>
<td>50</td>
<td>Between 40 &amp; 60</td>
<td>Between 30 &amp; 70</td>
<td>Outside of 30 &amp; 70</td>
</tr>
</tbody>
</table>

### Training List

- Raise the blue sliders to 55 collectively as a team staying within safe parameters so 30-70
- Raise one temperature up 10 and return to safe levels
☐ Lower one temperature down 20 and return to safe levels

☐ Raise one pressure up 10 and return to safe levels

☐ Lower one pressure down to 5 and return to safe levels

☐ Replace coolant

☐ Replace heating

☐ Replace catalyst

*Figure 5. Safety Indexes of the Simulation and Training List Provided to Participants*
Appendix F

Plant Functions Poster

Figure 6. Plant function poster
REFERENCES


