8-2007

# PHYSIOLOGICAL COMPLIANCE DURING TEAM PERFORMANCE

Amanda Elkins
*Clemson University*, aelkins@clemson.edu

# PHYSIOLOGICAL COMPLIANCE DURING 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
Amanda Nicole Elkins
August 2007

---

Accepted by:
Dr. Eric R. Muth, Committee Chair
Dr. Fred Switzer
Dr. Adam Hoover

ABSTRACT

Physiological compliance (PC) refers to the correlation between physiological measures of team members over time. The first goal of the current analyses was to generate several means to measure PC from heart rate variability (HRV) data. A second goal was to examine the relationship between PC and team performance during a building clearing task performed by 4-man teams. Teams were tasked with entering and clearing both real and simulated rooms populated with combatants (individuals with a weapon) and non-combatants (individuals without a weapon). Teams had to eliminate (shoot with a laser tag or simulated weapon) combatants and identify non-combatants (verbally or with a joystick). In Analysis I, linear correlation and directional agreement were shown to be the most sensitive PC measures when combined with HRV data. For Analysis II, 10 teams (20 subjects total, all male) were split into low and high performance groups based on their average team velocity and percentage of non-combatants acknowledged. Multivariate tests revealed a statistically significant difference between high a low performers, indicating that high, or better, performing teams tend to have higher PC. In Analysis III, one team was chosen to examine the relationship of performance and PC over time. Correlation testing on HRV data revealed a significant positive relationship between correlation RSA and performance ($r$=.853) and between correlation $\log_e$ RSA and performance ($r$=.859). These results suggest that PC may have merit for predicting team performance in a dynamic task. However, further research is needed.

## ACKNOWLEDGMENTS

I would like to gratefully acknowledge the help received from numerous individuals in completing this work. First and foremost, I would like to thank Dr. Eric Muth for his supervision, guidance, motivation and patience even during the slow start to this research. I will always remember that the body is a volume conductor. I would also like to thank Dr. Fred Switzer and Dr. Adam Hoover for their help during this work. I am indebted to Alex Walker and several other fellow graduate students that this would not have been possible without. Thank you all.

TABLE OF CONTENTS

Table of Contents (Continued)

Table of Contents (Continued)

LIST OF TABLES

LIST OF FIGURES

Figure Page

# CHAPTER 1
# INTRODUCTION

The growing complexity of technology has increased the difficulty of many workplace tasks previously completed individually (Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000). Many workplaces have begun focusing on teams and team centered approaches to work in order to meet this challenge. Effective training is essential for these teams to function successfully. While psychophysiological measures are often used in the field of human factors to evaluate the effect of work environments and training on people, these measures have often only been analyzed at the individual or dyad level (Henning, Boucsein, & Gil, 2001; Cacciopo & Petty, 1983). Physiological compliance involves taking a different approach by focusing on psychophysiological measures at the team level. Understanding how physiological compliance relates to team performance could lead to improved team training and assessment.

The first purpose of the current work was to generate a means to measure physiological compliance. In order to accomplish this, objective measures of physiological compliance scores derived from heart rate variability were created and explored. The distributions of various physiological compliance scores were also examined to confirm variability within the scores. The second purpose was to examine the relationship between physiological compliance and performance using the derived compliance scores.

## Teams and Training

A team can be defined as a set of two or more people who engage in cooperative, interdependent action in order to reach a common objective (Stout, Salas, & Fowlkes, 1997). The use of teams in the workplace has become increasingly widespread and can only be expected to become more prevalent in the future (Lucius & Kuhnert, 1997; Andrews 2005). Collaboration between workers is often used to improve performance and quality as well as to stay productive and competitive in target markets (Reich, 1987). For instance, the military makes use of teams regularly to accomplish a wide variety of mission objectives that could not be completed by a single person or multiple people acting alone.

While the use of teams has the potential to increase productivity and achievement within the workplace, the quality of the work to be completed by one of these teams is contingent on the quality of the team itself (Lucius & Kuhnert, 1997). Several factors may contribute to the success or failure of a team to perform effectively and efficiently (i.e. with high quality). For example, Chance (1989) contends that using personality as a selection criterion is important in quality teams because people with "team personalities" will be able to work together more effectively. Salas, Sims, and Burke (2005) contend that team leadership, mutual performance, backup behavior, team orientation, and adaptability are the most important factors in determining a team's successful performance. Most importantly, for the interests of the present analyses, training is also widely regarded as one of the key components in the quality of team performance.

A proper training procedure can enable teams to perform duties in an effective, efficient, and safe manner. The potential outcomes of having a good training procedure are attractive to most organizations, which in turn leads to interest and funding to training programs. For example, billions of dollars are spent each year on team training and necessary training support equipment in the military (Salas, Milham, & Bowers, 2003). Since a large portion of resources is devoted to improving training methods and procedures, organizations are both financially and emotionally invested in their chosen training methods. From an objective stand point, it seems both important and necessary for organizations to be able to conduct systematic evaluations of their training. Unfortunately, there is a lack of understanding of how to conduct evaluations of the training exercises at a team level and what the evaluations should entail (Stout, Salas, & Fowlkes, 1997).

Researchers are only just beginning to understand what comprises effective team training and this is in part due to the fact that studies frequently investigate individual behaviors in a team situation rather than focusing on team-level behavioral change resulting from training (Salas & Cannon-Bowers, 1997). At a team level, a more suitable analysis technique may be one that evaluates the individual in relation to the team, rather than just as an individual.

Evaluations of individuals at a team level are complicated by a number of factors including the work environment, the wide range of tasks to be completed by teams, and the diverse variables of interest in team performance (Stout, Salas, & Fowlkes, 1997). Despite these issues in evaluation, the military is one of the foremost consumers of training, and therefore, must have a way to identify and evaluate the efficacy of its training methods by assessing team performance.

While previous studies have suggested trainee reaction questionnaires, multiple choice knowledge tests, and performance assessment based on skill dimensions as methods for team training evaluation (Stout, Salas, & Fowlkes, 1997), it is possible that psychophysiological measures could provide some of the needed assessment capabilities in evaluation. Recent studies suggest a link between team performance and joint team physiological changes (Henning, et al, 2001; Henning, Ferris, & Armstead, 2006). If psychophysiological measures can provide an objective measure of team performance, the measure could then be applied to assessing training effectiveness.

## Psychophysiology

Psychophysiology is the branch of psychology devoted to studying the relationship between physiological and psychological functions and phenomena. Psychophysiological measures have often been used in studies of human-environment systems and team work (Henning et al, 2001). These measures have several advantages in the evaluation of environments and systems (Backs & Boucsein, 2000). First, psychophysiological measures are noninvasive. Second, psychophysiological measures can be acquired without requiring any explicit behavior from the subject and can be obtained unobtrusively. Lastly, they provide continuous information about an individual and are sensitive to physiological state changes.

The advantages of psychophysiological measures have led to their use in a variety of applied problems. For example, Springer, et al. (1990) used several psychophysiological measures to assess the benefits of computer aided design (CAD) systems when compared to a conventional drawing-board. While subjects performed a design task, heart rate and heart rate variability (HRV) were recorded. They found that using psychophysiological methods provided "hints" for design improvements of the CAD systems. The highest areas of strain, as determined by increased heart rate and decreased HRV, were revealed to be reading the design task, evaluating different solutions, and calculating. The authors were then able to focus on designs to minimize these stress areas.

Similarly, heart rate measures were used by Wientjes, ter Maat, and Gaillard (1994) to study workload and stress among a number of air traffic controllers. Heart rate showed an increase when the number of aircraft under control increased or when potential conflicts arose. It was concluded that variations in psychophysiology corresponded with variations in stress levels due to workload.

Unfortunately, previous research has mostly focused on the individual physiology and has failed to examine physiology at a group or team level (Henning et al, 2001). Even when multiple subjects have been studied, psychophysiological responses were often analyzed separately for each participant (Cacioppo & Petty, 1983). When Springer, et al. (1990) analyzed physiological data from the study on CAD systems, it was evaluated at the individual level. Similarly, Wientjes, et al. (1994) also looked at psychophysiological stress measures in individual air traffic controllers. While the majority of psychophysiological research continues to focus on individuals, the essential role of teams in the workplace has become indisputable (Stout, Salas, & Fowlkes, 1997). This gap only further highlights the need for more research examining physiology at a team level. Physiological compliance could serve to close the gap between research and the needs of team assessment in the real world.

## Physiological Compliance

Focusing on physiological compliance is one way to understand and apply psychophysiology at a group or team level. Physiological compliance refers to psychophysiological changes of a joint nature (i.e. 2 or more people) (Smith & Smith, 1987). Physiological changes that involve two or more people and exhibit close correspondence of reflected mutual influence are considered to be compliant (Henning & Korbelak, 2005).

Levenson and Gottman (1983) explained physiological linkage as physiological responses between members of an interacting dyad that "show considerable relatedness or linkage." In general, physiological compliance can be operationally defined as the correlation between physiological measures of team members over time.

Although physiological compliance is not a new notion in research, it is somewhat understudied (Henning, et al, 2001). A limited number of studies have departed the conventional individual level of psychophysiological data analysis and have explored the implications of psychophysiology among dyads. These studies present evidence that psychophysiological responses among interacting people can exhibit relatedness and linkage that can then be used to predict performance.

DiMascio, Boyd, Greenblatt, and Solomon (1955) conducted a study involving the physiological responses of psychotherapists and their clients during interviews. They found that heart rates of the therapists and client "often varied together." Later, in a study of the effects of a therapist's praise on female responses, Malmo, Boag, and Smith (1957) found more evidence of physiological linkage between people. In their study, the amplitude of the electromyogram of the examiner and patient both fell subsequent to the examiner issuing praise and remained constant after the examiner expressed criticism. The authors noted the need for objective investigation of the interaction between patient and therapist. Malmo, et al. (1957) clearly demonstrated the covariation of physiological signals of two people, i.e., physiological compliance, after each manipulation.

The psychophysiological relationship between individuals in dyads was also explored by Kaplan, Burch, and Bloom (1964). They examined skin conductance among dyads of participants who either "liked" or "disliked" each other. In this study, data from dyads composed of people that disliked each other were more likely to show significant physiological correlations compared to data from dyads composed of people who liked each other.

More recently, Levenson and Gottman (1983) looked at social interactions during topical discussions between thirty married couples and found that 60% of the variance in marital satisfaction was accounted for by physiological linkage. They found evidence that distressed couples' interactions were more likely to exhibit physiological interrelatedness and this linkage was more likely to occur in periods of high negative affect. They concluded that physiological compliance was associated with periods of negative affect and could therefore be problematic.

Hatfield, Cacioppo, and Rapson (1994) offered a different interpretation of Levenson and Gottman's findings. They suggested that physiological compliance is not problematic, but simply accompanies periods of intense social interaction.

In one of the most recent studies focused on physiological compliance, Henning et al. (2001) used two-person teams participating in a team tracking task to evaluate if physiological compliance is a determinant of team performance. They found that increased physiological compliance was associated with improved performance and decreased team tracking error. Furthermore, Henning et al. also found that there was no correlation of joystick control actions and physiological compliance, which countered interpretations that physiological compliance may have been attributed to "matched task behaviors resulting in matched physiological responses."

In a subsequent study, Henning and Korbelak (2005) studied the predictive value of physiological compliance in team performance. In this study, teams of two performed a tracking task randomly interspersed with unexpected switches in task control dynamics. The study revealed that higher cardiac physiological compliance predicted lower tracking error from teams, confirming the hypothesis that physiological compliance could be used as a predictor of team performance and possibly as a means of determining a team's preparedness to manage the unanticipated.

In perhaps the only study of the physiological compliance of a four-person team, Henning, Ferris, and Armstead (2006) analyzed HRV data and subjective teamwork effectiveness responses from a research team during meetings throughout a 6 month period. Although physiological compliance scores from entire meetings provided no predictive information, the data did show that physiological compliance during periods of sequential speech activity was predictive of team effectiveness ratings. Further study on this targeted approach was suggested.

Recent research has clearly demonstrated that physiological compliance has potential for assessing and predicting team performance. It has been shown to correlate with increased monitored performance as well as with subjective measures of team performance. There are several theoretical models that explain the predictive value of physiological compliance.

## Behavioral Cybernetic Model

Smith and Smith (1987) developed a cybernetic model of behavior that can guide the direction of studies in physiological compliance (Henning et al, 2001). This model looks at behavior as a closed-loop process where behavioral links are established through sensory feedback mediated by motor behavior.

Smith and Smith's model states that motor behavior serves as the principal means for individuals to self-regulate internal body states by using the direct effects of motor behavior on physiological states such as heart rate and hormone activity (Henning et al, 2001). By putting internal states of the body under the voluntary control of the individual, this model leads to a dynamic view of behavior in which individuals depend on internal and external sensory feedback from motor activity to self-regulate behavior (Smith & Smith, 1966). Unlike other models, physiological feedback is the creation point, not endpoint, of behavior.

In this model, social interaction is considered to be the closed-loop process of motor-sensory interactions between individuals that share control over a certain behavior (Smith & Smith, 1966). Motor-sensory interaction between two people leads to reciprocal changes in the internal states of both people. Figure 1 is an illustration of the joint regulation of behavior among two people. Motor sensory control by one individual in response to the behavior of the other individual serves as a source of sensory feedback to the other individual that is used to adjust the behavior, and so on. The ensuing behavior is influenced by the effects of motor-sensory interaction on internal physiological states.

This can be illustrated by the simple example of two people trying to lift an object together. Both people will likely watch the actions of the other person for information to determine which action to perform (i.e. lifting, lowering, pushing, etc.). This sensory feedback from motor actions would be considered social interaction which would affect each person's internal physiological state (i.e. heart rate increase in preparation for lifting). The following behavior (lifting) would be affected by the physiological state changes caused by the previous synchronized motor actions.

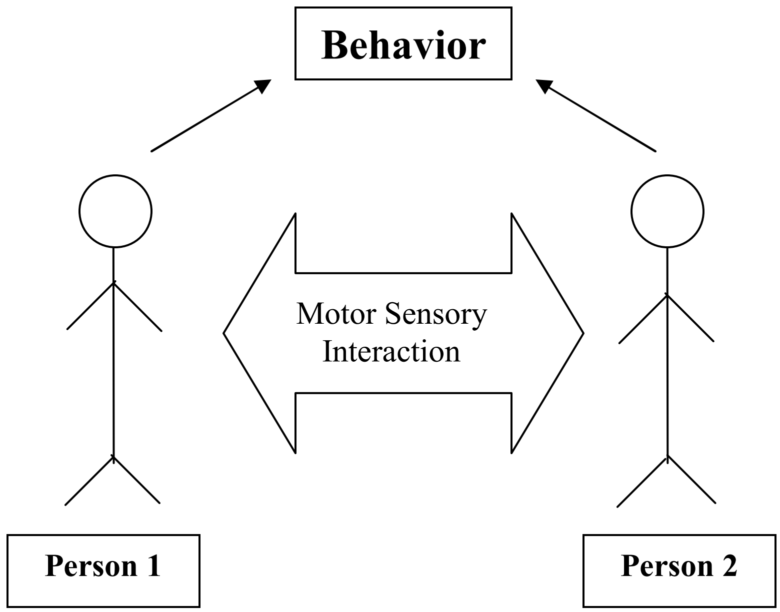



Figure 1. Social Interaction according to Smith and Smith's (1994) Cybernetic Model

An important distinction in this model is that physiological changes during social interaction must be "considered more than just a response to ongoing social behaviors" (Henning, et al, 2001). This distinction clarifies why joint physiological changes are considered physiological compliance and not response covariation. The general idea of Smith and associates' cybernetic model of behavior is that physiology drives feedback in the form of motor control and behavior.

Previous studies have demonstrated the importance of feedback on team performance and studies involving social tracking dynamics have already ascertained the important influence of feedback on team performance (Sauter & Smith, 1971; Kao & Smith, 1971). According to the cybernetic model, physiology drives behavior and social feedback; therefore, if the internal physiology of team participants is compliant, the social feedback between participants will be compliant and increase team performance.

## Mental Models

While the behavioral cybernetic model mentioned above focuses on physiological compliance as the beginning rather than end point of compliant behavior, another possibility is that physiological compliance is the endpoint of a process beginning with team mental models.

Mental models are structured information frameworks that allow individuals to describe and predict behavior (Norman, 1983). A team mental model can be defined as a shared understanding and mental representation of knowledge about relevant aspects of the team's environment (Mohammed & Dumville, 2001). Teams can have several mental models which can include tasks, roles, goals, and abilities.

Previous research has indicated that similar team mental models among teammates may increase the performance and effectiveness of team tasks (Beng-Chong & Klein, 2006). In 2006, Beng-Chong and Klein examined performance and mental model measures from 71 combat teams in the Singapore Army. Their results demonstrated that performance evaluations from subject matter experts were positively correlated with team mental model similarity, indicating that a shared mental model among team members was beneficial to team performance. Another study by Mathieu, Heffner, Goodwin, Cannon-Bowers, and Salas (2005) had comparable findings when examining taskwork mental model similarity of dyad teams on a flight simulator. The results revealed a positive correlation between mental model sharedness and performance measures.

This effect on performance is perhaps due to the ability of team members to accurately anticipate responses of other members with similar mental models, which allows them to coordinate successfully (Mathieu, et al, 2000). In essence, having similar mental models allows team members to prepare for and execute synchronous actions. Because of the similarity and relatedness of the actions, this preparation or execution affects the internal physiology of both members in the same way. The correspondence between the physiological variables of the team members can be considered physiological compliance.

Although this model does not assume that physiological compliance necessarily precedes compliant behavioral action, it does not devalue the possibility of using physiological compliance as a measurement of team training performance and predictor of future team performance. Quantifying the degree of physiological compliance between team members can still be related to training performance measures even if it is not the beginning point of action. From this view, physiological compliance would still be useful as an objective training metric.

## Physiological Covariation

Another view suggests that physiological compliance may be nothing more than covariation resulting from being exposed to the same environment. For example, it has been shown that RSA typically decreases in individuals under stress (Grossman, 1983). This would be true for more than one person under stress as well. If both people were exposed to the same stressor at the same time, it follows that they would both have decreases in RSA that might appear as physiological compliance.

If physiological compliance is actually covariation due to the environment, it is possible that people who display the most compliant physiology are experiencing the immediate environment in the most similar way. For example, if a group of people were startled, members who were actually scared would have increases in heart rate while members who were not would presumably have less of a change in heart rate. The similarity in experiencing the environment could be beneficial if it allowed those involved to focus on the same goals and actions, thereby improving their interactions. Even if physiological compliance is a byproduct of environmental experience, its value in objectively quantifying team performance remains intact.

## Physiological Measures

Physiological measures used to study physiological compliance in the past have included heart rate, skin conductance, pulse transmission time to the finger, and respiration (Henning et al., 2001; Levenson & Gottman, 1983). In at least one previous study, it has been shown that heart rate variability (HRV) showed the strongest predictive relationship with performance (Henning et al., 2001). The current study used HRV, mean successive differences, and mean inter-beat-intervals as measures of physiological compliance.

### Heart Rate Variability and Respiratory Sinus Arrhythmia

The autonomic nervous system is the portion of the peripheral nervous system consisting of the sympathetic (SNS) and parasympathetic nervous system (PNS) branches. These reciprocal branches regulate most of the internal states of the body. The PNS interacts with the autonomic nervous system through the vagus nerve and is responsible for heart rate deceleration and pupil constriction, among other things (Kimball, 2005). HRV is often used as an indirect measure of PNS activity, in part because cardiac response through vagal activity is quicker than that of the SNS (Bernston et al, 1997).

HRV is a measurement of the naturally occurring changes in an individual's inter-beat-intervals (IBIs) over time (McCraty, Atkinson, & Tomasino, 2001). The top numbers in Figure 2 illustrate IBIs, or the time between the R waves of 2 successive heart beats. The bottom numbers represent heart rate in beats-per minute (BPM). IBIs are inversely related with heart rate which means IBIs increase as heart rate decreases and vice versa.

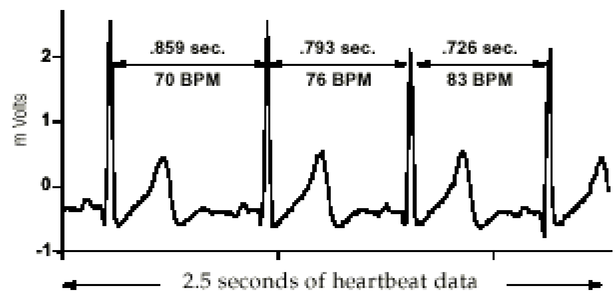



Figure 2. The relationship between heart rate and heart rate variability.

(McCraty, Atkinson, & Tomasino, 2001)

Variability in heart rate over time is predominantly composed of three frequencies. These frequencies are high, medium, and low. The component frequency of interest in the current analyses is the high frequency, which normally ranges from .15 Hz to .4 Hz. In the high frequency component, it can be shown that during exhale, IBI increases and during inhale, IBI decreases (Porges, 1995). This cyclical change is known as respiratory sinus arrhythmia (RSA) and, when respiration is controlled for, can be used as a measure of PNS activity. In the present study, high frequency HRV in the form of RSA was used as an objective physiological measure used for comparison across individuals. RSA was examined as raw RSA data and as $\log_e$ RSA due to the fact that raw RSA does not have a normal distribution.

## Mean Successive Difference

Mean successive difference (MSD) is a measure of the standard deviation of heart rate. The MSD statistic is computed as the average of the difference between successive IBIs for a particular time period (Allen, Matthews, & Kenyon, 2000). It filters out low frequency sources of variability in the IBI data series. MSD has been validated in previous studies and also has been shown to effectively track manipulated cardiac vagal control (Hayono, et al, 1991). In this study, MSD was explored as a possible measure to be used in measuring physiological compliance.

### Mean IBI

Mean IBI values are the average of a set of IBI data for a particular time period. Average IBI was examined due to its simplicity and direct relationship to a physiological system, i.e. rate of contraction of the heart, in contrast to the other measures that are indirectly related to PNS activity.

## Hypotheses

While previous studies have provided a great deal of information about physiological compliance, few, if any, have attempted the study of physiological compliance between team members performing a highly physical and applied task. The current analyses aimed to continue exploring physiological compliance as an objective assessment of team performance by examining compliance across individuals in four person teams as they perform building clearing tasks. The purpose of Analysis I was to examine several possible measures of physiological compliance using signal matching, instantaneous derivative matching, directional agreement, and correlation. Analysis I also sought to verify the existence of variability in these measures and systematic changes in this variability. Analysis II used the measures derived in Analysis I to examine the predictive correlation, if any, between physiological compliance scores during training and performance during testing.

Analysis III applied the physiological compliance scoring methods to one team to examine the relationship of performance and compliance over time.

The analyses described used data collected from 2 previous studies that took place at Clemson University and Clemson's Military Operations in Urban Terrain (MOUT) Shoot-House facility at the U.S. Army's 263rd Air & Missile Defense Command detachment in Anderson, SC.

Based on previous research mentioned above, it was anticipated that the current analyses would result in a usable physiological compliance score derived from HRV that would lead to an objective scoring method for team performance. Once the scores had been derived, it was also expected that:

1. There would be a greater degree of average physiological compliance during training among team members of a group who had "good" performance when compared to teams with "bad" performance during testing.
2. Physiological compliance would be positively correlated with performance improvements.

# CHAPTER 2
# GENERAL METHODS

## Data Set

### Participants

Participant data were mined from 2 previous studies on team training. For Analyses I and II, 10 teams of participants (40 participants total, all male, ages 18-30) were selected from a previous study entitled "Virtual Environment Training and Building Clearing." The previous study screened participants for good physical condition, moderate level of experience with first person shooter style video games, no formal combat or weapons training, and English as their first language (Carpenter, 2006). Participants were randomly assigned to 4-person teams and were compensated at the approximate rate of $10 an hour for their time.

For Analysis III, participant data were mined from a previous study on team training entitled "Establishing Team Training Metrics through the Use of a Virtual Training Lab." One team consisting of 4 male participants (ages 18-30) was selected based on performance improvement over time. The previous study screened participants for good physical condition, moderate level of experience with first person shooter style video games, no formal combat or weapons training, and English as their first language. Participants were compensated anywhere between $75 and $250 for their time.

### Task Description: Analyses I and II

As previously mentioned, participant data used in Analyses I and II were mined from a pre-existing "Virtual Environment Training and Building Clearing" study (Carpenter, 2006). There were four experimental conditions based on training in the study: video training, video and Xbox training, video and real world training, and video, real world, and Xbox training. The current analysis did not use data from the video training only condition. In the testing portion, participants completed six testing trials in the "shoot-house." Raw IBI data were collected during training and testing.

Participants took part in the training phase of the study in teams of four. Each team completed six trials that totaled about one hour of training time (Carpenter, 2006). Depending on the training method assigned, a trial was defined by either a predetermined amount of time (Xbox), or a number of room entries (real world). Subjects were given periodic feedback on their performance.

The real-world training consisted of subjects practicing the room clearing techniques by clearing either a square shaped room or an L-shaped room. Participants were armed with rubber M-16 replica rifles to enhance the training realism. Dummy targets were positioned prior to the clearing of each room. Teams were instructed to verbally acknowledge the paper targets as combatants (defined as holding a gun) or non-combatants (defined as not holding gun) upon entry into a room.

The four teams placed in the 'training video + Xbox training' condition received one hour of additional interactive, video game based training after viewing the training video. Teams performed room clearing exercises through a customizable 4-player mode on the *Rainbow Six 3: Black Arrow* video game for the XBOX video game console. The scenario used for training involved clearing a hotel, accessible on the *Mission* menu of the game. All subjects were set up to use only M-16 rifles in the game. All other weapons, thermal vision, and health upgrades were disabled. The experimenter stressed that the goal was not to carelessly rush through but to take the game seriously.

The four teams placed in the ‘training video + real-world training + Xbox’ condition received additional training consisting of a combination of 30 minutes of real-world training and 30 minutes of game-based Xbox training.

The testing task was building clearing while engaging combatants and recognizing non-combatants in a real-world environment. The teams were told to complete each trial while incurring a minimum number of casualties. They were to clear a five room shoot-house while attempting to engage combatants (individuals with a weapon) and acknowledge unarmed non-combatants (individuals without a weapon). Teams had to eliminate (shoot with a simulated weapon) combatants but only identify non-combatants. Performance metrics and heart rate data were collected during the task.

## Task Description: Analysis III

As previously mentioned, participant data for Analysis III were mined from a pre-existing study, “Establishing Team Training Metrics through the Use of a Virtual Training Lab.” Although several training conditions were used in this study, IBI data were collected during the testing phase only. Therefore, the data mined in this analysis were from the testing phase of the study.

The testing phase task was building clearing in a real-world environment while engaging combatants and recognizing non-combatants. The teams were to clear a five room shoot-house while attempting to engage combatants and acknowledge unarmed non-combatants. A map of the shoot-house can be seen in appendix A. Participants were told to complete each trial while incurring a minimum number of casualties. The number and location of combatants and non-combatants was varied throughout the trials. The number of non-combatants present ranged from 1 to 3, while the number of combatants present was either 1 or 2. The total number of combatant and non-combatants present never exceeded 4 during one trial. Performance metrics and heart rate data were recorded during the task. This study used twenty testing trials of the task in the real world to examine the success of the training techniques.

## Materials

The inter-beat-interval (IBI) data in the current analyses were obtained through the use of two physiological recording systems. During training session data collection at Clemson University, a UFI 3991x/1 –IBI Biolog (Morro Bay, CA) was used to acquire raw IBI data. This device recorded raw IBI data for off-line download and viewing. Physiological data collected at the Clemson University shoot-house were obtained by a UFI Wearable Arousal Meter v. 2.4a (WAM) (Morro Bay, CA). The WAM is wireless and derives IBIs from an amplified and filtered electrocardiogram (ECG) signal from three electrodes placed on the participant. It automatically detects and corrects errors in the IBI series. The WAM then transmits over an 802.11 link to a target computer where the data can be viewed.

Both devices used three disposable electrodes connected to the system through a Fetrode Input Assembly. One electrode was placed in the middle of the sternum and another was placed just below the ribcage on the left hand side of the body. The last electrode was placed on the lower right abdomen and was used as a reference point in order to filter out unnecessary noise in the recordings. This configuration was intended to allow for maximum detection of sequential electrical events of the cardiac cycle. Figure 3 illustrates a comparable electrode arrangement.

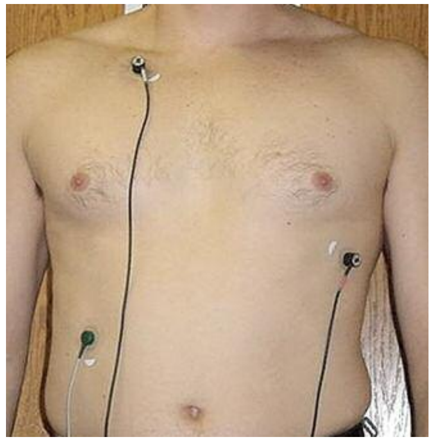


Figure 3. Electrode Configuration (Hoover & Muth, 2004)

## Data Processing

### Signal Enhancement

The devices mentioned above provide raw IBI data that must be prepared to be analyzed. In order to do this, several steps were taken to maximize signal strength and put the data in an analyzable format. Because of the presence of artifacts due to motion, telemetry, and noise, it was necessary to examine the quality of the raw data.

Initially, data files were examined and labeled usable or unusable. Files with errors comprising more than approximately 10% of the recording were considered unusable. In the data set for Analyses I and II, several teams had three subjects of usable data while other teams had only two. In order to balance this, the two subject data files with the least amount of artifact were chosen to be analyzed from teams with three usable team members. After this initial inspection, a total of 20 data files from the Analyses I and II data sets were left to carry into the next data processing step.

In the next step, the raw IBI data were divided or split based on training trials. To accomplish this, Biolog DPSx1.4 software (UFI, Morrow Bay, CA) was used to “region select” periods of interest in the raw data files and save each separately. For training data, periods of interest were defined by trials; therefore, the file of each participant from the Analyses I and II data sets were divided and saved as 6 separate files, with one file for each trial. For testing data used in Analysis III, the files were already separated by trial and no splitting was necessary.

Next, the split raw data were cleaned and artifacts were manually corrected with a locally developed editor program. This was true for both the uncorrected data from the Biolog and the autocorrected data from the WAM. Data used for Analyses I and II required an average of 6 corrections per 100 IBI values. The most common errors in the IBI recordings were detecting a false peak, which created a short IBI, and missing detection of a peak, which created a long IBI. The corrections used were “combine” (merge 2 consecutive IBI values), “combine 3” (merge 3 consecutive IBI values), “combine split” (merge 2 consecutive IBI values and split 2 ways), “combine 3 split 2” (merge 3 consecutive IBI values and split 2 ways), “combine 2 split 3” (merge 2 consecutive IBI values and split 3 ways), and “combine 3 split 3” (merge 3 consecutive IBI values and split 3 ways).

Figure 4 illustrates an IBI recording with several artifacts likely caused by participant motion. Figure 5 illustrates the same recording after several of the corrective methods mentioned above have been applied. Proceeding from left to right, the corrections used here were “combine 3,” “combine 3 split 2,” “combine split,” “combine 3,” and “combine.”


Figure 4. IBI data with artifacts.

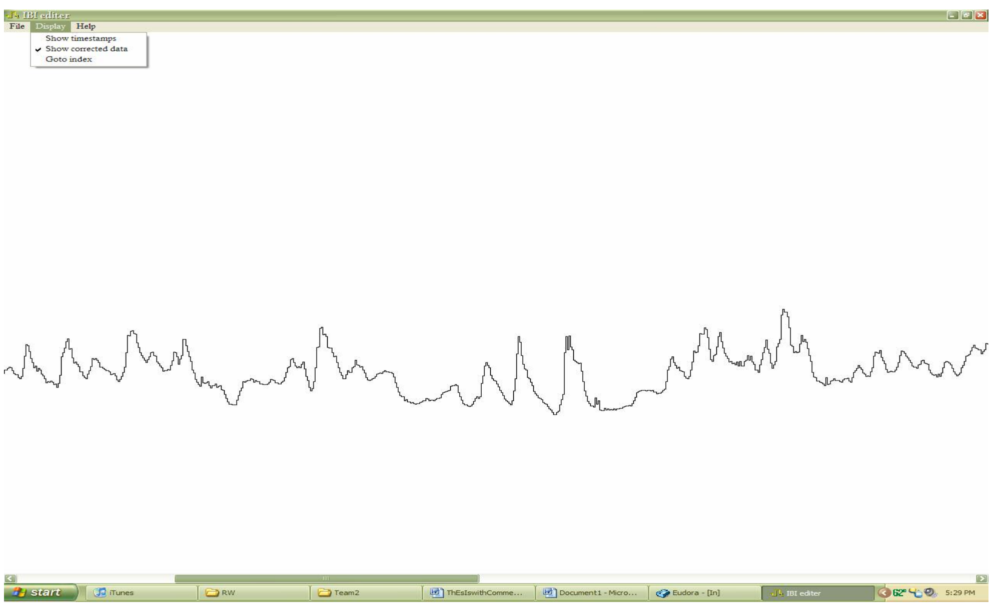



Figure 5. Cleaned IBI data without artifacts.

## Data Reduction

After data cleaning was completed, the data files from the Biolog were resampled through another locally developed program in order to create a continuous time series. This was not necessary for the data from the WAM as they were already synchronized. After resampling, each data file (one for each trial of each participant) was further split into 65 second windows. Sixty-five seconds were chosen as the window size because it was the smallest window usable by the program used in the next step. In the final data reduction step, each re-split data file was analyzed using additional locally developed software to gather statistics from them. Mean IBI values, mean successive differences, and the peak RSA frequency were all provided for analysis. RSA was derived using a fast Fourier transform which requires a power of two sample (in this case a 64 data point window) and the program skips the first data point (hence the 65 second windows). The windows did not overlap. The IBI data were analyzed by running spectral analysis using the 65 second periods of data. Spectral density estimates were derived at a bin width of 0.017 Hz [1 cycle per minute (cpm)]. The spectral power at the high frequency peak between 0.15 and 0.5 Hz was taken as the measure of RSA.

The same data reduction procedures outlined above were carried out, with some changes, for Analysis III data when the team to be used for analysis was chosen. The data provided by the reduction process were analyzed in Analyses II and II with the compliance methods developed in Analysis I.

## Performance Measures

For Analyses II and III, performance ratings were derived from a unit weighted linear model that included average team velocity and percentage of non-combatants acknowledged. Z-scores were derived for all data used in the model to guarantee equal weighting among the two scores since there was no evidence to suggest one was more important than the other. These measures were chosen from metrics that were monitored during the testing phase of the task because they model the speed-accuracy tradeoff present in human movement.

Team velocity was used as the speed portion of the performance rating and refers to the average speed of the team as a unit as the team members move through the environment. In preliminary analyses, average team velocity appears to be correlated with team performance, with expert teams tending to have a higher average velocity compared to novice teams. Subject matter experts have also stated that higher average team velocity is indicative of better team performance. In order to derive team velocity, velocity for each individual was computed at 20Hz intervals by computing the average distance the person moved in the surrounding half-second. This smoothing overcomes oversampling and noise in the data. The team velocity is the average of these 4 numbers for that time step. The average team velocity is then the average of these numbers throughout the recording. All 4 team members (rather than only the 2 selected for data analysis) are included in team velocity calculations in order to smooth the velocity. Once a person is shot (dead), his data are no longer included in the calculations.

Percentage of non-combatant acknowledgements was used as the accuracy portion of the performance rating and refers to the number of non-combatants acknowledged by a team relative to the number encountered by that team. For Analyses I and II, there were 12 possible non-combatants to be acknowledged over all testing trials. In Analysis III, there were a total of 36 possible non-combatants to be acknowledged. However, since some teams were unable to complete all trials successfully, it was also possible to encounter less than the total number of non-combatants present. To remedy this, the percentage of correct acknowledgements out of the total number of non-combatants confronted was taken. For example, if a team encountered 11 non-combatants and acknowledged only 8, they would receive a percentage of 72.7%.

This combination of performance metrics was chosen after examining the performance of each team on each measure individually. It was noted that teams were not consistently rated at the same position within both metrics. For example, the team with the highest average velocity, which would have indicated the best performing team, did not have the highest percentage of combatants acknowledged. This suggested that a metric that took the tradeoff between time and accuracy measures into account would be a more adequate performance metric. It is well known that a tradeoff exists between how fast a task can be performed and how many mistakes are made in performing the task. Usually, task performers can complete the task either very quickly with more errors or more slowly with fewer errors. Average team velocity and percentage of combatant acknowledgements seemed to quantify both ends of performance. Average team velocity represents "violence of action," or the aggressiveness used when entering a room for clearing. Aggressiveness is considered to be an asset when performing room clearing. On the other hand, the percentage of non-combatants is representative of carefulness used in correctly identifying the threat posed by people encountered during room clearing. An ideal team would use a balance of both by entering the room swiftly and aggressively but still managing to correctly identify non-combatants before firing. A correlation test ($r$=.09) confirmed that these two measures were related but not redundant, and were therefore good candidates for providing a performance score.

Other possible metrics included total number of team member casualties, number of completed missions, average time to complete (TTC), and test scores on a multiple choice knowledge tests. However, total number of casualties and number of completed missions were not used in the current performance model due to a lack of variability. For example, all teams in Analyses I and II completed 5 or 6 trials and several teams had the same number of casualties. TTC was not used because correlation testing showed that it was inversely redundant of average team velocity ($r$=-.720). Test scores on the multiple choice knowledge test were not included in the performance model after correlations revealed a negative correlation between test scores and average team velocity ($r$=-.679) and a negative correlation between test scores and non-combat acknowledgements ($r$=-.075), indicating that the test did not measure actual physical performance.

# CHAPTER 3
# ANALYSIS I

The purpose of the first analysis was to create several separate measures of physiological compliance through the analysis of mean IBI, mean successive differences, raw RSA, and $\log_e$ RSA data. The measures chosen were weighted scaled displaced matching (SM), instantaneous derivative matching (IDM), directional agreement (DA), and correlation. These measures provided 4 different ways of comparing the similarities of the curves of 2 team members.

## Method

### Signal Matching

SM was used to quantify physiological compliance by examining the differences in area between the data curves of team members. Figure 6 illustrates this concept. The area of interest between the curves is highlighted with lined shading. Greater area between the curves indicates less similarity between the signals, while less area between the curves indicates more similarity; therefore, a lower score on SM indicates higher physiological compliance.

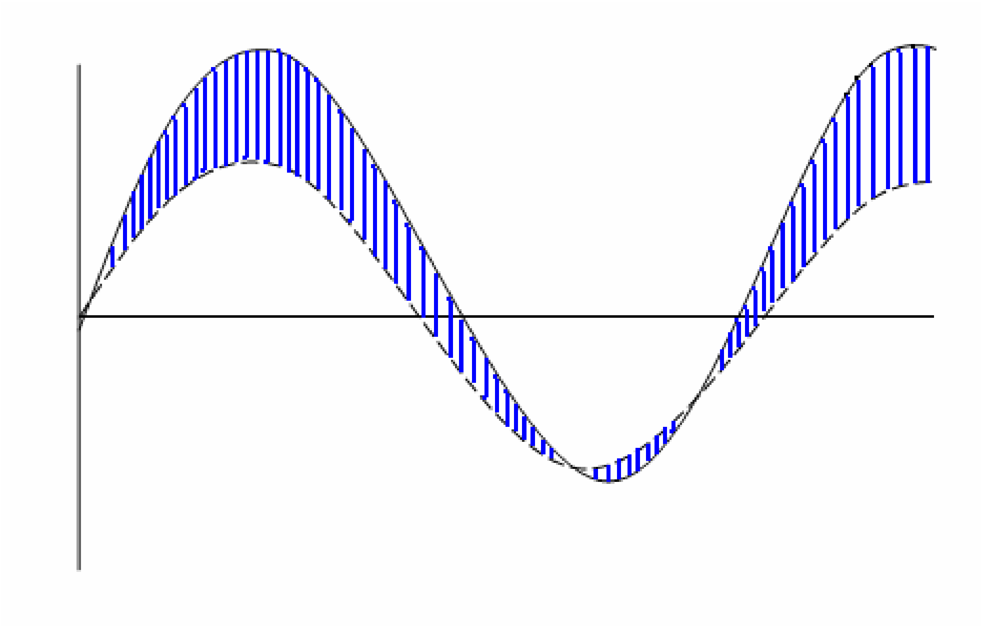


Figure 6. Example area between two data curves examined by signal matching.

The SM process was accomplished in several steps. First, the values from each physiological signal were normalized so that both signals were on an equal and comparable scale. To accomplish this, z-scores were derived for each data point in the data set. Next, the differences in each data point and its counter in the other signal were derived. For example, the difference between point 1 on team member A data would be compared to point 1 on team member B data, point 2 on team member A data would be compared to point 2 on team member 2 data, and so on. After differences had been derived, the overall mean of each team was used for comparison in Analysis II and the mean differences of each trial were compared in Analysis III. A lower mean indicated better physiological compliance.

Instantaneous Derivative Matching

IDM examined how well the slopes of two different physiological signal curves matched each other. The derivative of a point is the tangent to the curve at that point, which provides a slope. Because these analyses used discrete values, the slope for each point was calculated by the difference between that point and the next to get the instantaneous derivative for each point. The instantaneous derivatives for each point were then compared to the corresponding point on the signal from the other team member and the differences between each point were averaged. This is expressed by the following equation where *a* is team member A, *b* is team member B, and *t* is time:

$$\frac{1}{T}\sum_{t=0}^{T-1}\left|\left(a_{t+1}-a_{t}\right)-\left(b_{t+1}-b_{t}\right)\right|$$

Figure 7 illustrates an example of physiological signals from 2 team members. Four corresponding points on each show where instantaneous derivatives (represented by tangent arrows) would be compared to the corresponding derivative on the other curve. These curves would have a low score due to the similarity of the derivatives at each point, indicating high physiological compliance.

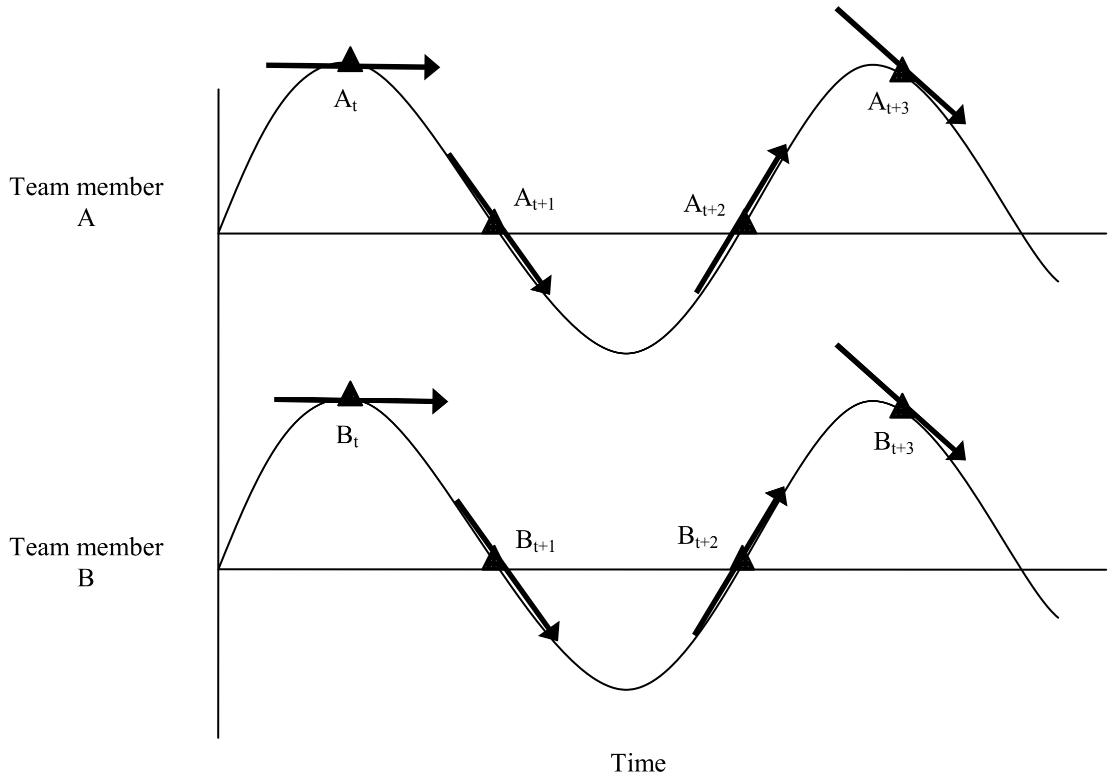



Figure 7. Instantaneous derivatives of 2 signal curves.

## Directional Agreement

DA provided a very basic measure of physiological compliance. The directional movement of each data point relative to the previous point was determined. For example, if the value at data point 1 is lower than the value at data point 2, the directional movement would be labeled as "increasing," but if the value at data point 1 is higher than the value at data point 2, the directional movement would be labeled as "decreasing."

Next, both team members' data were compared and determined if they were in directional agreement at each data point (i.e. both are increasing or both are decreasing). Figure 8 illustrates this concept with example data. In this figure, points $A_{t+1}$ and $B_{t+1}$ are in directional agreement because both are increasing from the previous point. However, points $A_{t+3}$ and $B_{t+3}$ are not in agreement because one increases relative to the previous point while the other decreases.  Points $A_{t+5}$ and $B_{t+5}$ both increase relative to the previous point and are in directional agreement.

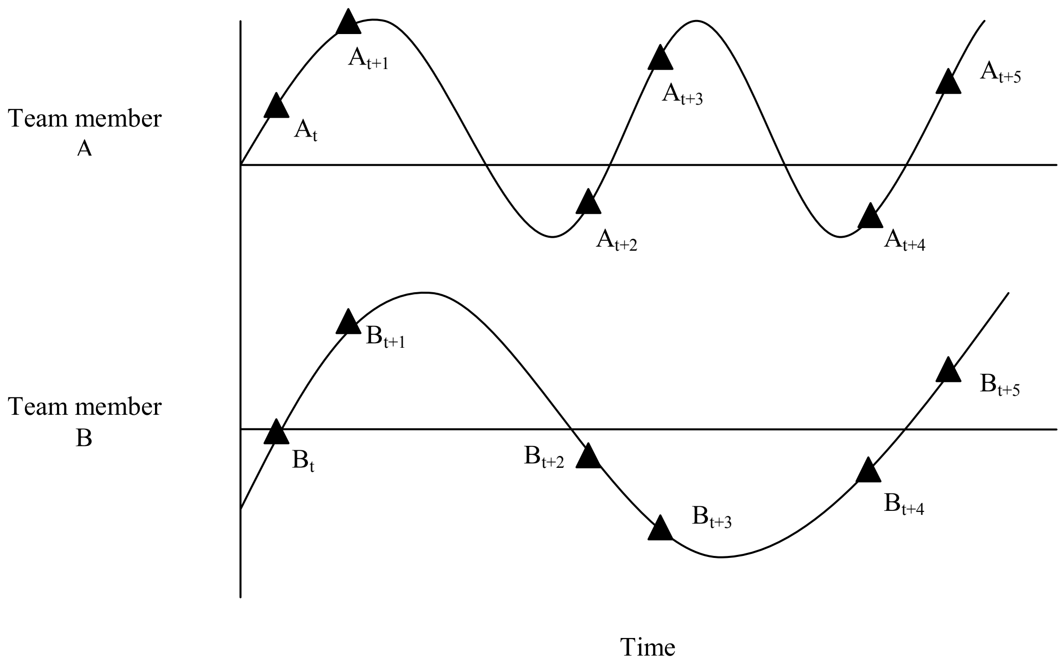



Figure 8. Illustration of directional agreement/disagreement.

A percent agreement was derived from the comparison explained above and used as a measure of physiological compliance.  For example, using only the 3 points mentioned previously, the percent agreement would be 66%.  Higher percentages represented higher levels of physiological compliance.

## Correlation

A correlation was used to indicate the strength of the linear relationship between the team members being examined.  This is possible because the current data have been synched over time.  For Analyses I and II, correlation coefficients were calculated over all trials to provide one score for the degree to which physiological compliance was demonstrated between team members. For Analysis III, correlation coefficients were calculated 2 ways. First, cleaned non-split IBI data were combined into 4 trial groupings and a correlation value was obtained for each grouping. Second, split IBI data were used to obtain mean IBI, RSA and $\log_e$ RSA values. Once again, these values were combined into 4 trial groupings and a correlation value was obtained for each grouping. Positive correlation values were related to physiological compliance with higher positive correlation values indicating higher levels of compliance.

## Criteria for a Measure of Compliance

In order to ensure that the measures described above were accurate descriptors of physiological compliance, several tests were run. First, a descriptive variability test was performed on all values derived from the measures. The purpose of this was to ensure that there was in fact variability in the data that could be measured. For example, if the correlation measure provided physiological compliance values of .9 for all teams, the lack of variability in the measure would prevent any relationship to variables of interest, such as performance, from being measured. The descriptive check was also used to ensure that any present variability was behaving in an explainable and expected way so that the data set was usable for further analysis. For example, if a data set was expected to have normal distribution but instead had a positively skewed distribution, the measure might lack face validity.

A visual check was also used along with the descriptive variability check. After data reduction, plots of mean IBI, MSD, RSA, and log RSA data for each team were visually scanned by 3 raters to look for the appearance of physiological compliance. Raters were instructed to rate each plot by determining if each trial exhibited compliance and by determining if the plot exhibited overall compliance. The definition of physiological compliance was explained to raters; however, no specific strategies or rules to be used in the rating process were given. The descriptions of strategies as reported by each rater are available in appendix B. When all raters did not agree, the rating with majority vote (2 out of 3) was chosen. Comparing the actual physiological compliance scores with visual ratings served to partially verify that the measures mentioned above actually gauged physiological compliance trends. The visual check was also used as a validity check to ensure that the values were varying in the "correct" direction to be sure the chosen measure is quantifying what it supposed to (physiological compliance). In other words, the check ensured that higher physiological compliance values were associated with teams that exhibit similar trends in the visual check.

## Results

### Variability Check

The descriptive variability check revealed sufficient variability within the scores obtained from all measures. However, normal distributions of data did not always indicate suitable variability. Some measures were expected to behave differently, such as being positively skewed rather than normal. Histograms showing the variability of each combination of physiological compliance and physiological measure can be seen in appendix C.

Table 1 shows the descriptive variability data obtained by using SM for each physiological measure. MSD and $\log_e$ RSA values showed normal data distributions; however, mean IBI and RSA values both showed positively skewed data sets. The differences of the distributions are explained by the scoring process. During the process to obtain SM scores, all values were converted to z-scores before the absolute difference between each point was taken. SM displayed normally distributed scores for all measures before the absolute value of each difference was taken; however, taking the absolute value sometimes changed the appearance of the distribution curve. For example, a distribution centered at zero would be expected to be positively skewed once all negative values were transformed to positive because of the symmetric nature of the distribution the negative values simply fold over onto the positive values. Essentially the negative portion of the curve is removed, leaving only the positive portion of the curve. However, distributions centered around a positive or negative value would be expected to still look normally distributed after taking the absolute value of all points. For instance, if all the values were centered at -1 before taking the absolute value, they would be centered at +1 afterwards, creating a normal distribution. Before taking absolute values, mean IBI and raw RSA distributions were normally centered at zero and therefore appear skewed in the final data set. MSD and $\log_e$ RSA distributions were centered about a positive value and displayed a normal distribution. It was also noted that the variability of each measure was fairly similar, which was expected due to the fact that all scores were derived from z-scores. Overall, SM provided adequate and explainable variability with each physiological measure.

Table 1. SM descriptive validity data.

| | Mean IBI | MSD | RSA | logRSA |
|---|---|---|---|---|
| **mean** | 0.935 | 1.0172 | 0.7916 | 1.0377 |
| **SD** | 0.67906 | 0.57438 | 0.86036 | 0.26225 |
| **shape** | positively skewed | normal | positively skewed | normal |
| **lowest value** | 0.29 | 0.24 | 0.22 | 0.67 |
| **lower quartile** | 0.3759 | 0.24 | 0.338 | 0.8144 |
| **median** | 0.6831 | 0.9964 | 0.4338 | 1.0188 |
| **upper quartile** | 1.3894 | 1.4895 | 0.9433 | 1.2681 |
| **highest value** | 2.38 | 2.02 | 3.12 | 1.51 |
| **behavior explainable** | Yes | yes | yes | yes |

Table 2 illustrates descriptive variability data for physiological data using IDM. All physiological measures showed positively skewed data distributions. Once again, this is explained due to the absolute values used with IDM. Score distributions for mean IBI, MSD, raw RSA, and $\log_e$ RSA data were normally distributed about zero before absolute values of the slope differences were taken, therefore, each of these displayed positively skewed distributions after scores were obtain with IDM. As expected, the variability of each measure using IDM differed because each measure was on a different scale. The variability of each scale also behaved as expected. Both mean IBI and RSA continued to show the largest variance, as with the other physiological compliance methods. $\text{Log}_e$ RSA also continued to have the least variability due to its small scale size. The data from IDM was found to be consistent with expected behavior.

Table 2. IDM descriptive variability data.

| | Mean IBI | MSD | RSA | logRSA |
|---|---|---|---|---|
| **mean** | 38.7667 | 7.5495 | 20.0848 | 1.2291 |
| **SD** | 9.66548 | 2.50686 | 18.3329 | 0.3992 |
| **shape** | positively skewed | positively skewed | positively skewed | positively skewed |
| **lowest value** | 27.57 | 5.77 | 6.07 | 0.79 |
| **lower quartile** | 30.9332 | 5.77 | 8.1424 | 0.8964 |
| **median** | 36.6225 | 6.3964 | 13.539 | 1.1564 |
| **upper quartile** | 45.7548 | 9.2283 | 24.1517 | 1.4186 |
| **highest value** | 55.98 | 13.5 | 67.89 | 2.1 |
| **behavior explainable** | Yes | yes | yes | yes |

Table 3 lists the descriptive variability data for directional agreement. Once two extreme values had been removed from RSA and $\log_e$ RSA data, all measures used with directional agreement provided normally distributed data sets. As anticipated, the values were normally distributed about 50%. It was expected that the values would be normally distributed about 50% because that is equal to chance. Even data with no compliance could achieve values near 50%, causing a grouping around that area. Each measure used with directional matching had expected similar variability since each was on the same percentage scale. Directional matching was determined to provide consistent data with variability.

Table 3. Directional matching descriptive variability data.

| | Mean IBI | MSD | RSA | logRSA |
|---|---|---|---|---|
| **mean** | 0.588 | 0.5127 | 0.5054 | 0.5001 |
| **SD** | 0.14294 | 0.07146 | 0.11364 | 0.11308 |
| **shape** | Normal | normal | normal | normal |
| **lowest value** | 0.43 | 0.4 | 0.27 | 0.24 |
| **lower quartile** | 0.473 | 0.4 | 0.4526 | 0.456 |
| **median** | 0.5495 | 0.4942 | 0.4783 | 0.4891 |
| **upper quartile** | 0.7093 | 0.5809 | 0.5976 | 0.5789 |
| **highest value** | 0.88 | 0.63 | 0.69 | 0.64 |
| **behavior explainable** | Yes | yes | yes | yes |

Table 4 illustrates descriptive variability data for the correlation measure. Although the means for RSA and $\log_e$ RSA data seem to be low when compared to mean IBI and MSD data, this is actually due to the ranges being distributed fairly evenly around zero. The negative and positive values cancel each other so that the mean is close to zero. Each physiological measure provided a normal data output when paired with the correlation method. It was also noticeable that mean IBI provided the only correlation distribution that did not go below zero. It is likely that this was due to the fact that IBIs of people involved in the same physical activity are likely positively correlated to some degree. For example, if 2 people decided to begin running at the same time, both will experience a decrease in mean IBI values. This was also a likely contributor to the smaller range for mean IBI data since the values could not go below zero. Other than the lack of negative values for mean IBI data, correlation measures all provided the expected similar variability due to the common scale between them. It was determined that using a simple linear correlation provided data variability and consistency.

Table 4. Correlation descriptive variability data.

| | Mean IBI | MSD | RSA | logRSA |
|---|---|---|---|---|
| **mean** | 0.2812 | 0.1436 | 0.0231 | 0.0314 |
| **SD** | 0.19188 | 0.17163 | 0.22391 | 0.21551 |
| **shape** | Normal | normal | normal | normal |
| **lowest value** | 0.09 | -0.14 | -0.3 | -0.29 |
| **lower quartile** | 0.1538 | 0.0528 | -0.162 | -0.1021 |
| **median** | 0.2225 | 0.0751 | -0.0121 | -0.0351 |
| **upper quartile** | 0.3622 | 0.3594 | 0.2512 | 0.2429 |
| **highest value** | 0.72 | 0.38 | 0.39 | 0.34 |
| **behavior explainable** | Yes | yes | yes | yes |

## Visual Check

Independent t-test analyses were used to compare visual ratings with scores obtained from each combination of physiological measure and compliance method. This testing looked for significant differences between compliant and non-compliant teams (as assigned by visual ratings) that would suggest measures are measuring the occurrence of compliance and are sensitive enough to detect performance differences. Table 5 provides a summary table of combinations of physiological measures and compliance methods that showed statistical significance.

Table 5. Measures achieving statistical significance during visual and compliance score comparison.

| | mean IBI | MSD | RSA | $\log_e$ RSA |
|---|---|---|---|---|
| **SM** | no | yes* | no | no |
| **IDM** | no | no | no | no |
| **DA** | yes | no** | yes | yes |
| **Correlation** | yes | no** | yes | yes |

*marginally significant, p<0.10, in the wrong direction
**trend in the correct direction

When testing visual ratings and scores derived from using DA, independent t-tests revealed statistically significance differences between visually rated compliant and non-compliant teams for mean IBI, raw RSA, and $\log_e$ RSA data. When using mean IBI data, compliant teams (mean=68.82%, SD=13.62%) had a significant higher mean than non-compliant teams (mean=48.78%, SD=4.83%) teams, t(8)=-3.10, $p<0.05$. When using DA in conjunction with raw RSA data, testing confirmed a statistically significant difference between the means of the compliant (mean=61.14%, SD=5.4%) and non-compliant (mean=43.48%, SD=.8.06%) teams, t(8)=-3.802, $p<0.05$. Additional testing using $\log_e$ RSA data indicated a statistically significant difference between mean $\log_e$ RSA values of compliant (mean=57.14%, SD=8.50%) and non-compliant (mean=45.27%, SD=10.92%) teams, t(8)=-1.82, $p<0.05$. Although no statistical significance was found, using MSD data with DA still provided scores that trended in the correct direction with compliant teams (mean=53.75%, SD=9.55%) having higher means than non-compliant (mean=49.61%, SD=5.39%) teams, t(8)=-.88, $p>0.05$.

Additional independent t-tests also revealed statistically significant differences between correlation scores of compliant and non-compliant teams when using mean IBI, raw RSA and $\log_e$ RSA data. When using mean IBI data, a statistically significant difference was discovered between the correlation scores of compliant (mean=.39, SD=.21) and non-compliant (mean=.17, SD=.07) teams, t(8)=-2.13, $p<0.05$. A statistically significant difference was also found between mean correlation compliance scores of compliant (mean=.16, SD=.21) and non-compliant (mean=-.07, SD=.20) teams when using raw RSA data, t(8)=-1.8, $p<0.05$. Further independent t-tests using $\log_e$ RSA data provided statistical significance when using a linear correlation to attain compliance scores; compliant teams (mean=.22, SD=.17) once again displayed a statistically higher mean than non-compliant teams (mean=-.09, SD=.14), t(8)=-3.15, $p<0.05$. As with DA, using MSD data with correlation did not provide statistical significance, but did provide scores that trended in the correct direction. Compliant teams (mean=.21, SD=.18) had a higher mean than non-compliant teams (mean=.10, SD=.16), t(8)=-.99, $p>0.05$.

Apart from one exception, no other statistical significance was observed when comparing visual rankings and compliance methods. The exception noted occurred when a marginal statistically significant difference in the incorrect direction was found between compliant (mean=1.32, SD=.66) and non-compliant (mean=.82, SD=.46) teams when using MSD data in conjunction with SM, t(8)=-1.42, $p<0.10$.

It is also notable that while DA and correlation methods provided data trending in the correct direction even when lacking statistical significance, SM and IDM methods displayed inconsistent trends. When using SM, means for mean IBI and RSA data were lower for compliant teams relative to non-compliant teams (recall that with SM, lower scores indicate more compliance). However, SM displayed trending in the opposite direction when combined with MSD and $\log_e$ RSA data by indicating that means were lower for non-compliant teams relative to compliant teams. IDM methods also displayed contradictory trending with different physiological data sets. When combined with RSA and $\log_e$ RSA data, IDM showed that, as expected, compliant teams had a lower mean than non-compliant teams. Conversely, when paired with mean IBI and MSD data, IDM exhibited trending in the opposite direction by indicating that non-compliant teams had lower scores than compliant teams. Table 6 provides a summary of the correct or incorrect trends of each measure.

Table 6. Mean differences trending in the correct direction

| | **mean IBI** | **MSD** | **RSA** | **log Rsa** |
|---|---|---|---|---|
| **SM** | yes | no | yes | no |
| **IDM** | no | no | yes | yes |
| **DA** | yes* | yes* | yes | yes* |
| **Correlation** | yes* | yes* | yes | yes* |

*statistically significant, $p<0.05$

## Discussion

Although the preferred measure of physiological compliance is not yet clear cut, this analysis attempted to use SM, IDM, DA, and linear correlation as methods of quantifying the degree of physiological compliance present in team members. These methods have not been applied to analyzing teams in such a dynamic task in any known previous study. Analysis I focused on developing the methods and exploring their variability distributions.

All methods suggested physiological compliance measures passed the descriptive variability check and were determined to behave in the manner expected. However, some method and physiological measure combinations seemed to provide better variability than others. Mean IBI, RSA, and $\log_e$ RSA all provided high and similar variance for the DA and correlation methods, suggesting that those physiological measures provided a consistent story. However, using SM and IDM often provided very dissimilar variances between the physiological measures, implying inconsistencies in the values derived from those measures.

The purpose of the visual check was to verify that the derived measures were actually measuring physiological compliance by comparing compliance values associated with teams that exhibited like tendencies in the visual check. In doing this, the visual check provided further evidence that when used with DA and correlation, mean IBI, RSA, and $\log_e$ RSA all provide consistent measures. Each was found to indicate significant mean differences of visually rated compliant and non-compliant teams in compliance measures. However, the visual check also showed that SM and IDM provided inconsistent results, demonstrating that those measures would not likely be suitable measures for physiological compliance.

The commonality among measures determined to be unsuitable seems to be loss of directional information. When using IDM and SM for calculations, the absolute value of differences between points on each line is taken in order to measure the magnitude of the differences. On a similar note, when MSD was calculated by the locally developed software mentioned in data reduction, the absolute value of the differences from one point to the next are taken to once again measure magnitude without regard to direction. Both IDM and SM had skewed data sets due to using absolute values. While the MSD data distributions for DA and correlation were normal, they were slightly more positively skewed than the rest of the data sets for DA and correlation. It is possible that directional information is needed in measures of physiological compliance due to the fact that using only absolute values skews the data set and decreases sensitivity to possible compliance.

By taking both the variability check and visual check into account, it was decided that DA and correlation methods combined with mean IBI, RSA and $\log_e$ RSA provided reliable and valid measures of physiological compliance between team members. Since no evidence existed to suggest any one of these measures should not be used, all six were carried forward into Analyses II and III. In Analysis II, the physiological compliance scores were used to facilitate comparison of teams characterized as good or bad performers and to examine a correlation comparison of physiological compliance scores and performance scores. In Analysis III, physiological compliance scores allowed examination of the relationship between performance and physiological compliance over time (trial) in a selected team.

# CHAPTER 4
# ANALYSIS II

The purpose of the second analysis was to examine the relationship between performance of 4-man teams and physiological compliance as measured by the methods mentioned above. Performance was measured as a combination of average team velocity and percentage of non-combatants acknowledged during a testing session and statistical analyses were used to analyze the data.

## Method

### Design

The first part of Analysis II was a between-subjects design. The independent variable was performance, derived from the weighted linear model described previously. The 10 teams with 2 members each were split into high and low performance groups based on the performance scores. Because values had been standardized using z-score values, the split between performance groups was made at zero. This also ensured an equal number (5) of teams in each group. The dependent variable was physiological compliance measured by each of the six measures carried on from Analysis I.

The second part of Analysis II was a correlation test to examine the linear relationship between the physiological compliance scores and performance scores.

## Physiological Compliance

DA and linear correlation were used in conjunction with mean IBI, raw RSA, and $\log_e$ RSA data to provide 6 physiological compliance measures. The data from each team were run through each method in order to get an average physiological compliance score for each team. The scores were compared among like scores (i.e. DA-IBI scores were only compared with other DA-IBI scores, not DA-RSA, correlation-IBI, etc.).

## Data Analysis

There were several separate statistical analyses run in Analysis II. For the between subjects comparison, a multivariate ANOVA was run on the average physiological compliance scores of high versus low performance teams for each of the 6 compliance measures. A multivariate ANOVA was chosen in order to remedy the increased probability of error associated with running multiple t-tests. A linear correlation test was also used to examine the relationship of the performance measure and physiological compliance.

## Results

A multivariate ANOVA revealed a statistically significant difference between the mean physiological compliance scores of high (mean=.66, SD=.17) and low (mean=.51, SD=.05) performance groups when using the combination of DA and mean IBI data, $F(1,8)=3.48$, $p<0.05$, as well as a statistically significant difference between physiological compliance scores of high (mean=.16, SD=.19) and low (mean=-.10, SD=.16) performance groups when using correlation and $\log_e$ RSA data, $F(1,8)=5.31$, $p<0.05$. A marginally significant difference of compliance scores between high (mean=.55, SD=.08) and low (mean=.45, SD=.12) performance groups was also revealed when using DA and $\log_e$ RSA data, F(1,8)=$p$<0.10.   Although all measures trended in the expected direction, no other statistically significant differences were found between high and low performance groups. Figure 9 illustrates the means of physiological compliance for the high and low performance groups using the 3 correlation combinations while figure 10 illustrates the means of physiological compliance for high and low performance groups using the 3 DA combination methods.

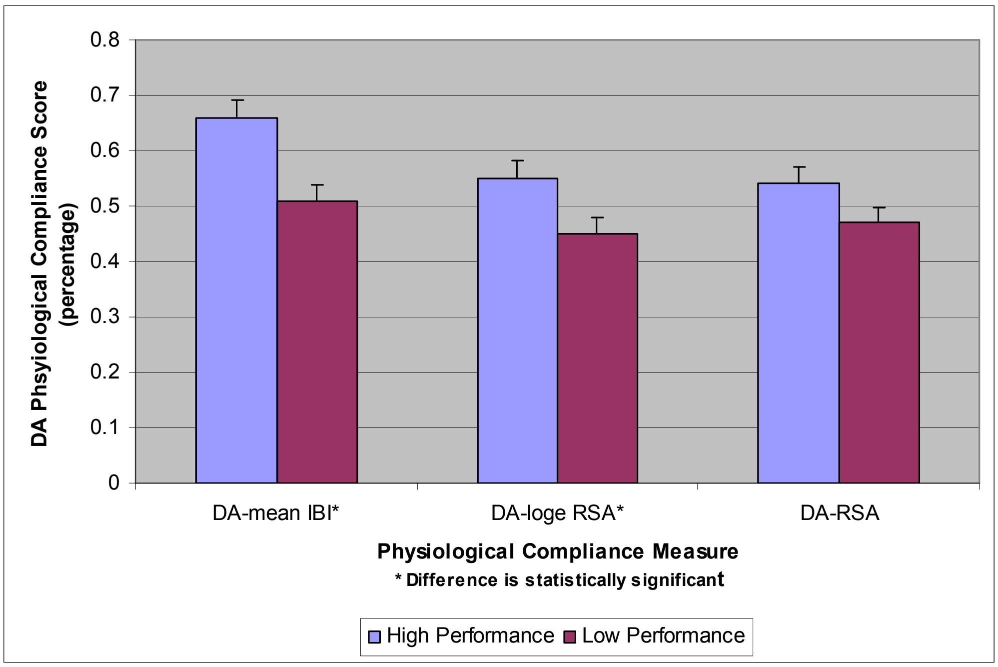



Figure 9. Mean physiological compliance for high and low performance teams by DA measure.

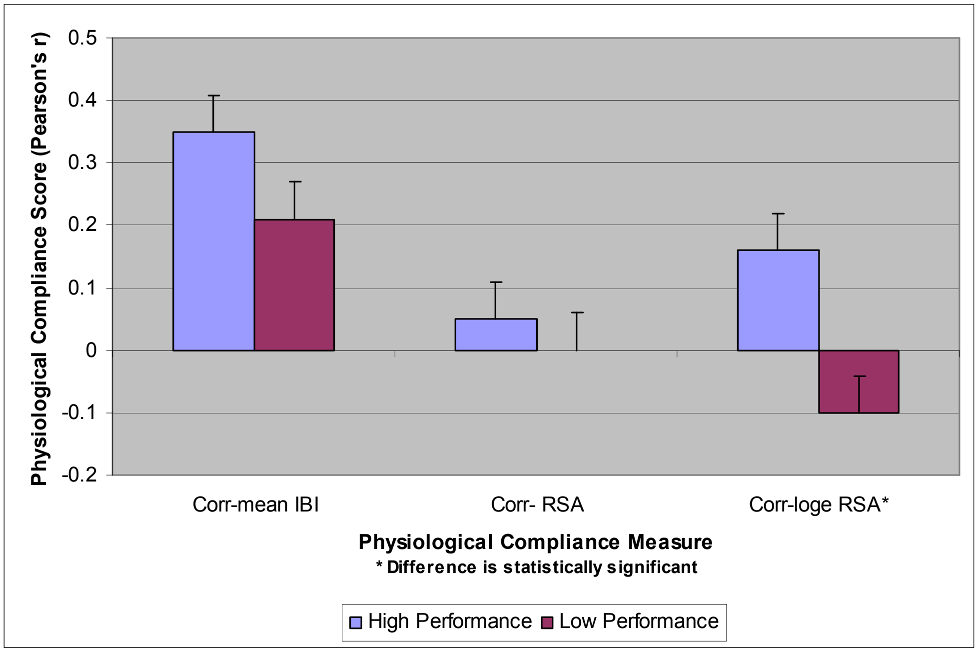



Figure 10. Mean physiological compliance for high and low performance teams by correlation measure.

A follow-up independent t-test was run on average team respiration rate (in cycles per minute) to rule it out as a confounding variable in the raw RSA and $\log_e$ RSA scores. An independent t-test revealed no statistically significant differences in respiration between high (mean=13.2, SD=2.07) and low (mean=12.34, SD=1.29) performance groups, $t(8)=-.791$, $p>0.05$.

Pearson correlation testing between all 6 measures and performance scores revealed no statistically significant relationship between physiological compliance and performance.

However significant correlations were found between DA-mean IBI and DA-$\log_e$ RSA, DA-mean IBI and correlation-mean IBI, DA-RSA and DA- $\log_e$ RSA, DA-RSA and correlation RSA, DA-RSA and correlation-$\log_e$ RSA, DA-$\log_e$ RSA and correlation RSA, DA-$\log_e$ RSA and correlation-$\log_e$ RSA, and correlation-RSA and correlation-$\log_e$ RSA. Table 7 shows the correlation values associated with each of the measures.

Table 7. Pearson's *r* correlation values for each physiological compliance measure and performance scores.

| | **DA-mean IBI** | **DA-RSA** | **DA-$\log_e$ RSA** | **Corr-mean IBI** | **Corr-RSA** | **Corr-$\log_e$ RSA** | **Performance** |
|---|---|---|---|---|---|---|---|
| **DA-mean IBI** | - | .401 | **.649*** | **.833*** | .326 | .392 | .058 |
| **DA-RSA** | .401 | - | **.907*** | .189 | **.624*** | **.591*** | .137 |
| **DA-$\log_e$ RSA** | **.649*** | **.907*** | - | .392 | **.562*** | **.607*** | .121 |
| **Corr-mean IBI** | **.833*** | .189 | .392 | - | .268 | .391 | -.133 |
| **Corr- RSA** | .326 | **.624*** | **.562*** | .268 | - | **.557*** | .130 |
| **Corr-$\log_e$ RSA** | .392 | **.591*** | **.607*** | .391 | **.557*** | - | .140 |
| **Performance** | .058 | .137 | .121 | -.133 | .130 | .140 | - |

*** Correlation is significant at the 0.05 level (1-tailed)**

## Discussion

The purpose of Analysis II was to begin to explore the relationship between performance and compliance. This analysis used 6 different methods to measure physiological compliance: DA and mean IBI data, DA and raw RSA data, DA and $\log_e$ RSA data, correlation and mean IBI data, correlation and, raw RSA data, as well as correlation and $\log_e$ RSA data.

Although the physiological compliance methods and performance scoring model used were exploratory in nature, the results, at least in part, suggest that there is a positive relationship between physiological compliance during training and performance during testing. The multivariate tests revealed a statistically significant difference between high a low performers, indicating that high, or better, performing teams tend to have a higher physiological compliance.

While the performance measures used in overall performance scoring have yet to be validated, it is believed that using average team velocity and percentage of non-combatants acknowledged allowed the performance metric to be sensitive to both ends of the performance spectrum (speed and caution). Along the same lines, it is believed that physiological compliance measures derived for and used in this analysis provided an adequate look at actual physiological compliance. The significant correlations between the measures served to reinforce that the measures were indeed measuring the same phenomenon, as they were intended to.

The results of the multivariate test suggests that using correlation with $\log_e$ RSA, DA with $\log_e$ RSA, and DA with mean IBI provide the most sensitive measures for detecting physiological compliance. However, all other measures also expressed means trending in the same direction, so it is possible that the lack of significance in those measures and in the correlation testing is due to the limitations of this analysis.

The main limitation of this analysis lies in the low number of subjects. Although data from 20 subjects were used, this only provided 10 physiological compliance scores and 10 performance scores for comparison. Standard deviations were often large, especially in correlation-$\log_e$ RSA calculations. By increasing the number of participants, the effect of variability could be lessened and the capacity to detect significant differences could be augmented by increasing statistical power.

Despite the limitations, this analysis was successful in exploring the relationship of physiological compliance and performance. It also provided additional evidence of a possible positive relationship and support for continued studies on this point. Analysis II aimed to do this by exploring the relationship of physiological compliance and performance over time.

# CHAPTER 5

# ANALYSIS III

The purpose of Analysis III was to examine the relationship of physiological compliance and team performance over time. As mentioned previously, the data were mined from a previous study. A team that illustrated improved performance over time was chosen.

## Methods

### Design

For Analysis III, a team that showed improvement over time was chosen for within-subjects analysis and comparison. The independent variable was time and the dependent variables included performance and physiological compliance.

## Data Processing

Due to the smaller data sets used in this analysis, several adjustments in data processing were made for this analysis.

### Signal Enhancement

The same signal enhancement process described in Analysis I was employed in this analysis with a few exceptions. Because data used in Analysis III were obtained through the WAM device that sampled at 4 Hz and all team members were saved in one file, the data had to be prepared for enhancement. Locally developed software was used to downsample the data to 1 Hz and to split each member into separate data files for the next step. Each trial was saved as a separate file, so there was no need to perform region select splitting as done previously.

In the next step, IBI data were cleaned in another locally developed software program, as described in analysis I. However, the WAM device used to gather data for Analysis III has built-in algorithms for auto correcting data, so there were far fewer data corrections to be made by hand. In this data set, only an average of 2 of every 100 IBIs needed to be corrected.

## Data Reduction

The resampling step mentioned in Analysis I was also not necessary in this analysis because the WAM outputs already synched IBI data. The data files were also split into 65 second windows for analysis. However, the data for each trial were also saved in cleaned IBI format for separate analysis.

As in Analyses I and II, re-split data files were analyzed using additional locally developed software to gather statistics from them. Mean IBI values and the peak RSA frequency were provided for analysis. RSA was also recalculated to obtain $\log_e$ RSA values.

## Performance

Although the chosen team completed 20 trials, the trials were split into 5 groups of 4 trials each (trials 1-4, 5-8, 9-12, 13-16, and 17-20) for analysis. This was done because of the nature of the performance metric that was developed in earlier analyses and carried over into this analysis. Because this metric took percent of non-combatants acknowledged into account, several trials needed to be lumped together in order to provide adequate variability within that portion of the score. For example, through one trial, there may have only been 1 non-combatant present, allowing for scores of only 0 or 100%. By using the percent acknowledged from 4 trials, the variability was increased considerably. The trials were grouped by a factor of 4 because teams had received feedback sessions after every 4 sessions when participating in the actual study.

The team chosen for the focus of this analysis was required to meet two criteria. The first was that it had to demonstrate an improvement in performance over time. This was examined by calculating performance scores for each team and plotting each on a line graph so that improvements would be visible. It also was required that the chosen team had adequate "usable" data. Due to the presence of artifacts in physiological data of highly mobile subjects, it was necessary to examine the data and ensure that the chosen team had at least two members that exhibited data with errors less than 10% of the time throughout at least 3 of every 4 trials. Figure 11 illustrates the performance improvement of the team chosen (out of 23) after meeting both criteria.

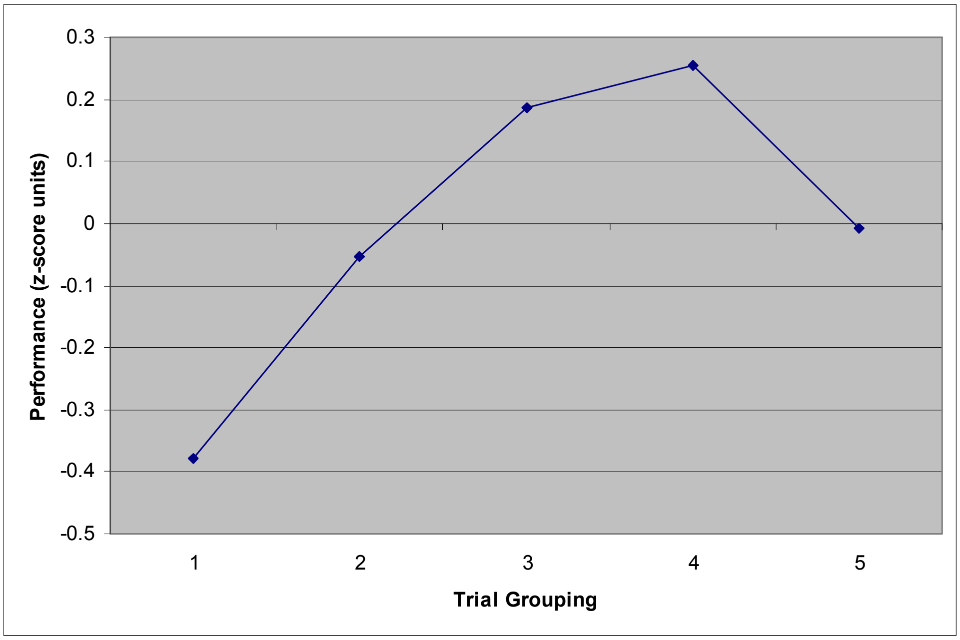



Figure 11. Performance scores by trial grouping for Analysis II team.

## Physiological Compliance

The shorter data files in this analysis, due to shorter trial times often shorter than 2 minutes, made it necessary to calculate physiological compliance in several different ways. Due to constraints of the performance metric, the data used were grouped together in 4 trial groupings.

First, calculations of correlation and DA were completed using non-split clean IBI data from each trial. Each set of IBI data for every 4 trials were stacked on each other to allow for a correlation to be computed. Correlations were then calculated for all IBIs from both team members in 4 trials. This method provided one score of DA and correlation of IBI data for each trial grouping. However, RSA and loge RSA were not able to be calculated by hand so compliance was measured in a second way.

As in Analyses I and II, split data files were used to obtain RSA values with locally developed software. However, the downfall of using this approach was that averages over

65 second windows were required and some trials were less than 130 seconds (2 windows). Therefore, only one RSA value was obtained for some trials. DA was not able to be used on data with only one RSA point (no point for directional comparison). Only correlation physiological compliance scores were calculated for mean IBI, RSA, and $\log_e$ RSA data from split files. These scores were also combined by grouping every 4 trials.

Overall, physiological compliance was measured five ways: correlation and DA from cleaned (non-split) IBI files and correlation of mean IBI, RSA, and $\log_e$ RSA from split data files.

## Data Analysis

Separate correlation tests were used to look for significant correlations between performance scores and physiological compliance scores derived from different measures. A repeated measures ANOVA was also employed to rule out covariation effects of respiration on RSA.

## Results

When examining correlation and DA on non-split IBI data, correlation tests revealed no statistically significant correlation between DA or correlation physiological compliance measures and performance.

Further correlation testing on split IBI data revealed a significant positive relationship between correlation RSA and performance ($r$=.853), as well as a statistically significant positive correlation between correlation $\log_e$ RSA and performance ($r$=.859). No statistically significant relationship was found between correlation IBI and performance.

Figure 12 illustrates the relationship between performance and RSA and $\log_e$ RSA correlation physiological compliance scores.

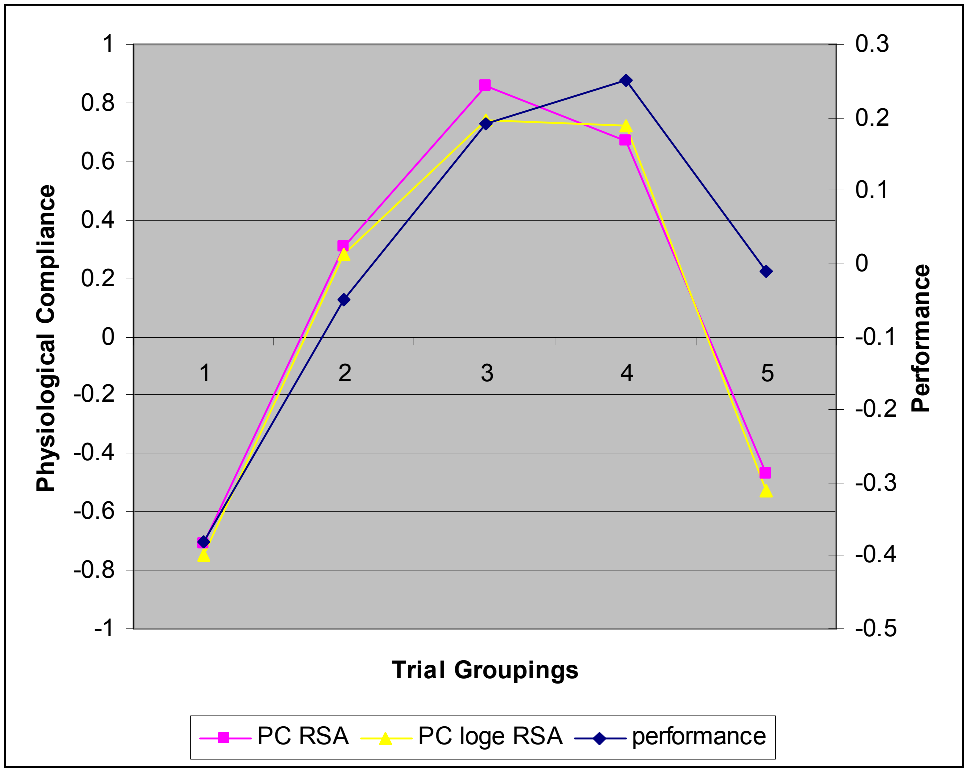



Figure 12. Relationship between performance and physiological compliance determined by correlation of RSA and $\log_e$ RSA values.

It is also important to note that a repeated measures ANOVA revealed no significant effect of respiration over time, therefore ruling out covariation effect possibilities, $F(4,20)=1.295$, $p>0.05$.

## Discussion

Analysis III aimed to examine the relationship of physiological compliance and performance over time by using several different methods to calculate physiological compliance. Two of the physiological compliance measures provided evidence that performance and physiological compliance are positively correlated, as stated in the hypotheses. Physiological compliance seems to increase incrementally with performance as performance improves.

Figure 12 also gives an illustration of the relationship between physiological compliance and performance even when performance does not increase. In the last trial grouping, the team acknowledged only half of the combatants encountered but also had their highest average team velocity overall. These performance values were located near the extremes and served to cancel each other out in the overall performance score (resulting in a score near 0). As performance scores decreased, physiological compliance decreased, further exemplifying the relationship between the two variables.

Three of the 5 measures used did not display a statistically significant relationship between compliance and performance. It is possible that correlation with mean IBI data from split files is not sensitive enough to detect a relationship, as it failed to in Analysis II as well. Furthermore, using non-split IBI data for measures probably resulted in a loss of significance due to the data itself. When using mean IBI values, the numbers are averaged and are therefore more stable. However, small differences between individual IBIs can change overall correlations when means are not used.

Analysis III brought several issues about measuring physiological compliance to light. It was noted that the length of the time period for physiological compliance to be measured has a great effect. Trial time periods in Analysis III were all mostly shorter than 2 minutes. This led to problems when trying to average over 65 second windows in order to get RSA values. This is a downfall that will be experienced any time an FFT is used to obtain values because it must average to derive these values. A similar weakness was shown in DA as a physiological compliance method. Like RSA values, DA requires more than one 65 second window of data in order to look at agreement for RSA or $\log_e$ RSA. The drawbacks of these measures indicate that this should be taken into account when designing a study to examine physiological compliance.

Another issue brought forth by Analysis III was the constraints of the performance measure chosen. While the measure itself seems to accurately quantify team performance, using percentage of non-combatants acknowledged was perhaps not the best for the data from this particular study. In order to use the measure, trials were grouped, decreasing the number of data points in the analyses from 20 to 5. Once again, this leads back to the indication that this should be taken into account when designing a study at the beginning.

Overall, the task used to get the data for Analysis III was not optimally designed to examine physiological compliance, which made analyzing it more difficult. When studying physiological compliance and physiology in general, longer data files are desirable for manipulation in the data reduction phases. However, Analysis III revealed a relationship between performance and compliance despite the shortcomings of the data set.

# CHAPTER 6
# GENERAL DISCUSSION

The objective of these analyses was to begin to explore the relationship of physiological compliance and performance by creating possible physiological compliance measures and applying them to existing data. Analysis I was successful in creating 6 viable compliance measures. Variability checks showed physiological compliance did vary among teams, allowing for statistical tests to be run in order to examine the differences. Subsequent analyses supported the hypotheses that a positive relationship between compliance and performance existed. These findings show that the compliance-performance correlation is not limited to stationary tasks and can be applied even in complex task settings.

These analyses revealed that DA and linear correlation combined with mean IBI, RSA, and $\log_e$ RSA data are all valid measures of physiological compliance. These all showed expected variability and provided results indicating that performance and physiological compliance are positively correlated (some statistically significant, some correctly trending). However, results from Analyses II and III also suggest that some of these measures are more sensitive to physiological compliance than others. DA seems to work the best when paired with mean IBI or $\log_e$ RSA data while linear correlation seems to work best when combined with RSA or $\log_e$ RSA data.

It is likely that RSA data provided sensitive measures for physiological compliance due to its quick response through vagal activity (Bernston et al, 1997). It is often used as an indirect measure of PNS activity for this reason. This is in agreement with the behavioral cybernetic model proposed by Smith and Smith (1966). They assert that physiological compliance occurs before the behavior; therefore, the mechanism chosen to measure it must react quickly in order to provide a good measure of compliance. It should be noted that the sensitivity of these measures could also be attributed to similar mental models among team members. Anticipation of the same events could also lead to measurable changes in these measures. However, these measures do suggest that physiological covariation is not the origin of physiological compliance. The teams examined in these analyses were completing a task that did not require a simultaneous start or action throughout the task. Team members often completed separate tasks in different areas (i.e. splitting off to clear separate parts of rooms). If physiological covariation were the basis of physiological compliance, it is likely that significance would not have been measureable without considerable phase shifting and including only parts where team members acted together.

The current analyses justified the physiological compliance measures by examining variability and visual checks, but other work should be done to further validate the measures. For example, more work is needed to prove the reliability of the findings mentioned here. It is imperative to the validity of these physiological compliance measures that the results are repeatable. Also, these measures should be used in a study designed specifically for their use in order to ensure that all possible confounds can be ruled out. This would allow for the relationship between performance and physiological compliance to be examined as performance decreases as well. Validating these measures would provide more firm evidence to link physiological compliance and performance.

These findings provide compelling evidence to suggest a positive relationship between compliance and team performance, which would mean that physiological compliance may be an important part of team proficiency. The overall results are in agreement with the most recent research in this area that also signifies a positive relationship

(Henning, et al, 2001). Henning et al. explored several measures of physiological compliance including cross correlation and weighted coherence. As in the current analyses, they found that the correlation measure showed the strongest predictive relationship with performance. It seems that the simplest measures used were shown to be the most sensitive measures. This suggests that measures devised to quantify physiological compliance should be straightforward and uncomplicated and include both direction and magnitude.

It is possible that in the future, measures of physiological compliance will find many uses throughout the field of human factors. As mentioned previously, physiological compliance could become an integral part of assessing team training methods and performance. It would provide a constant, objective assessment of performance that is sorely needed in the billion dollar area of training. Team members could be selected on the degree to which they exhibited physiological compliance while completing training. Physiological compliance could be an important design tool when designing systems for multiple simultaneous users. It would offer continuous monitoring by using physiological devices and objective scoring on prototypes. Cooperative work stations could be rated on the extent that they encourage physiological compliance (Henning, et al, 2001).

Although there are countless possible applications for using physiological compliance as an objective score of team performance, it could also be an indicator of negative performance. For example, some teams are formed for the express purpose of having different viewpoints and opinions and it is possible that physiological compliance among those team members would be detrimental to performance. An example of this would be a design group trying to formulate new ideas. Therefore, it must be noted that while physiological compliance can be a good measure of performance for many team activities, it will not apply to all situations. The relevance of physiological compliance should be assessed for each situation individually.

## Future Directions

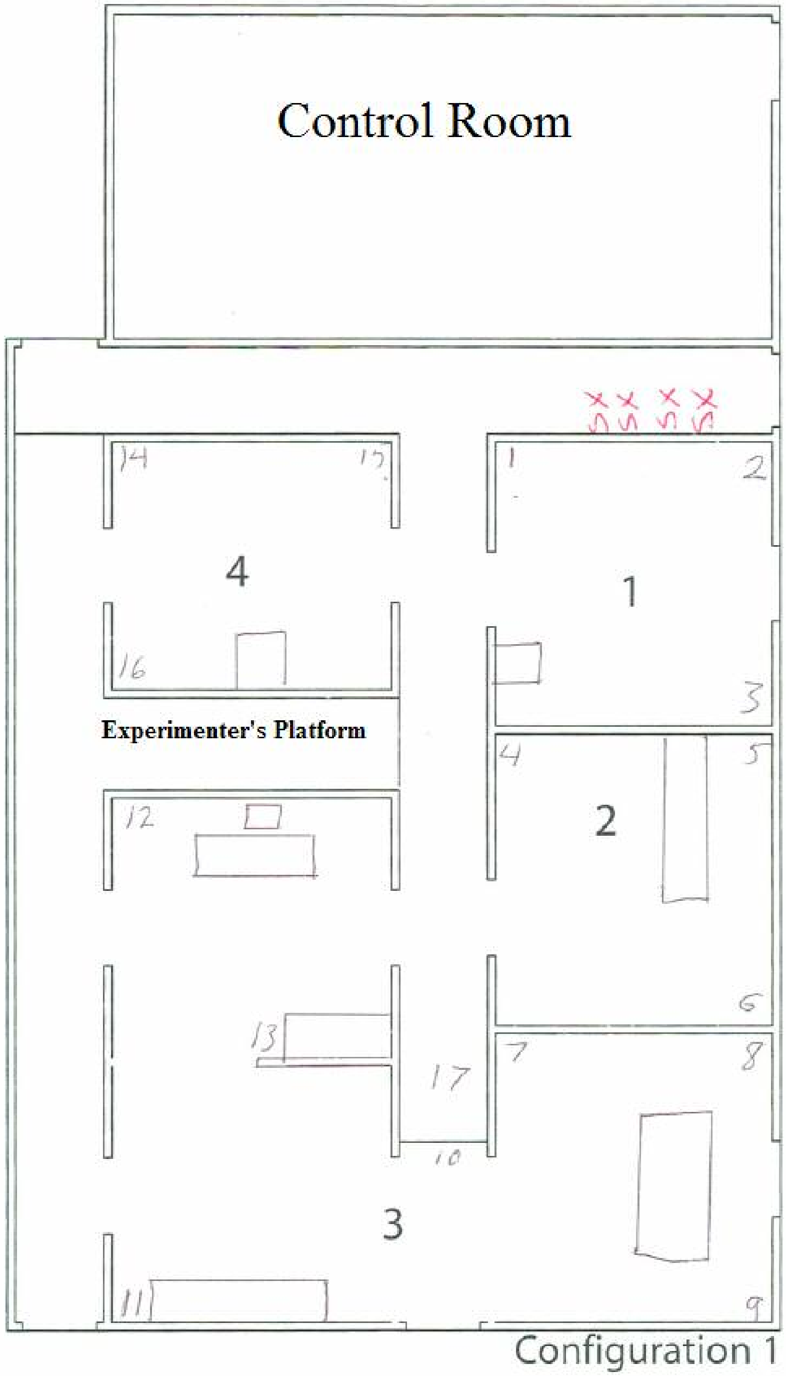
Since there is no "best way" as of now to measure physiological compliance, prospective studies should also continue exploring new possible measures of compliance. It is possible that physiological measures not used in the current analyses could provide a reliable measure of physiological compliance. Future studies should also seek to look at physiological compliance in new ways, such as in the workplace or classroom.

## Conclusion

Overall, these analyses achieved their objective of examining the possible linkage between physiological compliance and performance and found evidence supporting the small body of literature that already suggests an important positive relationship. Several measures such as correlation and DA used in combination with mean IBI, RSA, and $\log_e$ RSA data were shown to be sensitive to changes in physiological compliance. These measures merit further examination and validation in future studies. This study, combined with the results of existing studies, presents adequate evidence that physiological compliance could be a useful tool in training and numerous other fields and should continue to be explored in order to be fully understood.

APPENDICES

Appendix A:  Shoot-House Map



# Appendix B: Rater Descriptions

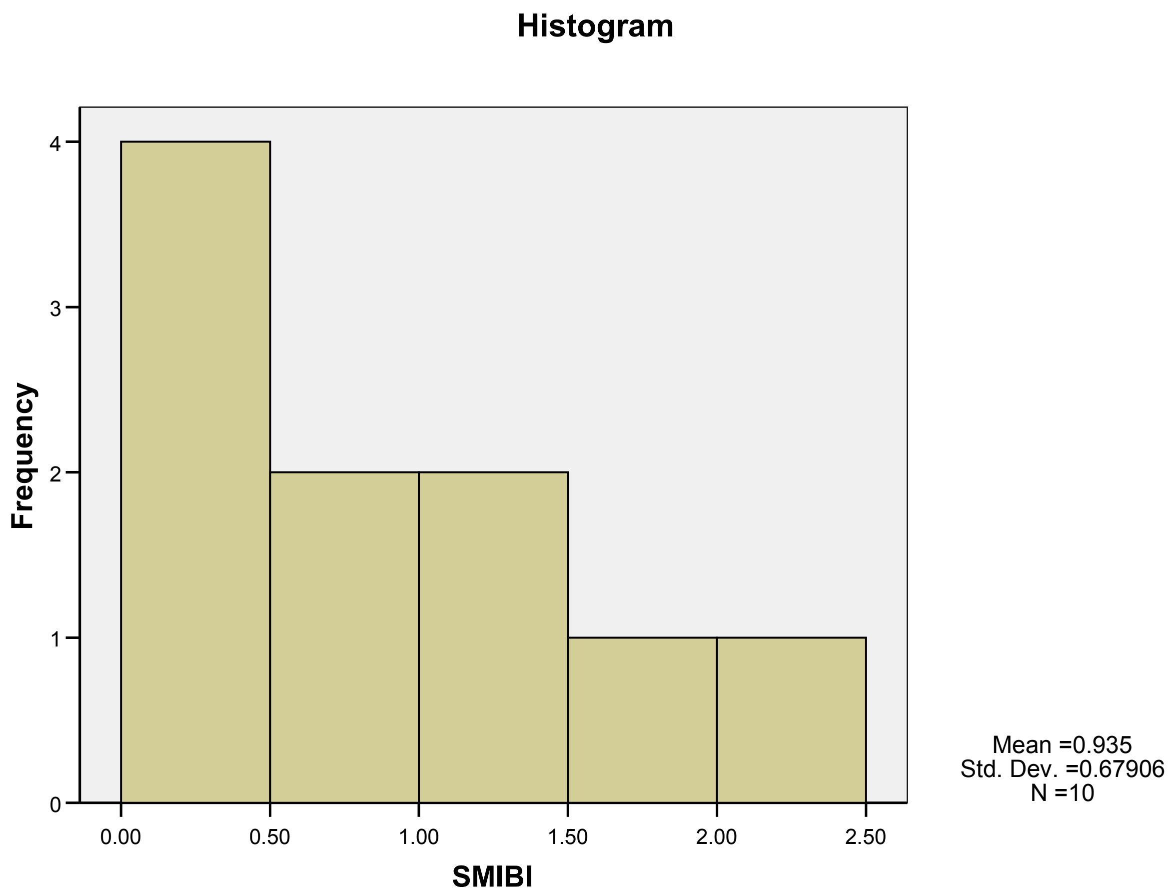
**Rater 1:** "I tried to see if the lines followed the same trend from one point to the next. I also looked to see if the 2 curves went go in the same direction (regardless of magnitude). If I was confused I used the one that matched the most points."

**Rater 2:** "I took a liberal approach to determine if the data were compliant. I looked at the peaks and valleys of the data and compared how many there were and where the occurred. Sometimes the peaks and valleys of one person lagged behind another, but I still considered them to be compliant. If 2 people were "sort of" compliant, I examined how much of the data corresponded. If more data corresponded than didn't, I considered them to be compliant. Sometimes I would hold the graphs out at arms length and purposely blur my eyes to see if I could see any sort of pattern."

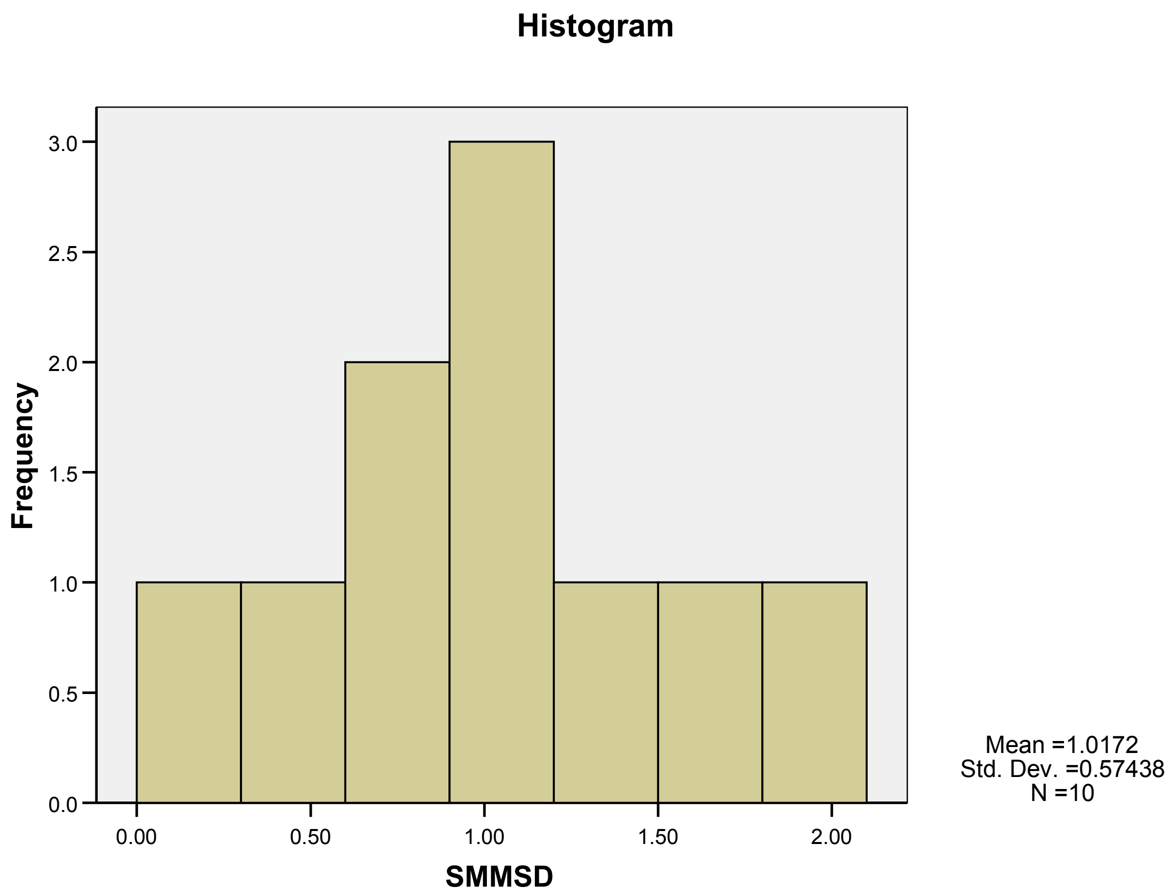
**Rater 3:** "I looked for a combination of things. I looked to see the lines went in the same direction, not considering the scale. I also looked for a subjective 'feeling' of compliance by looking at the lines overall (disregarding trial separations) to see if I felt like they expressed the same overall trends of up and down, even if not exactly at the same moments.

Appendix C: Histograms of distributions from Analysis I

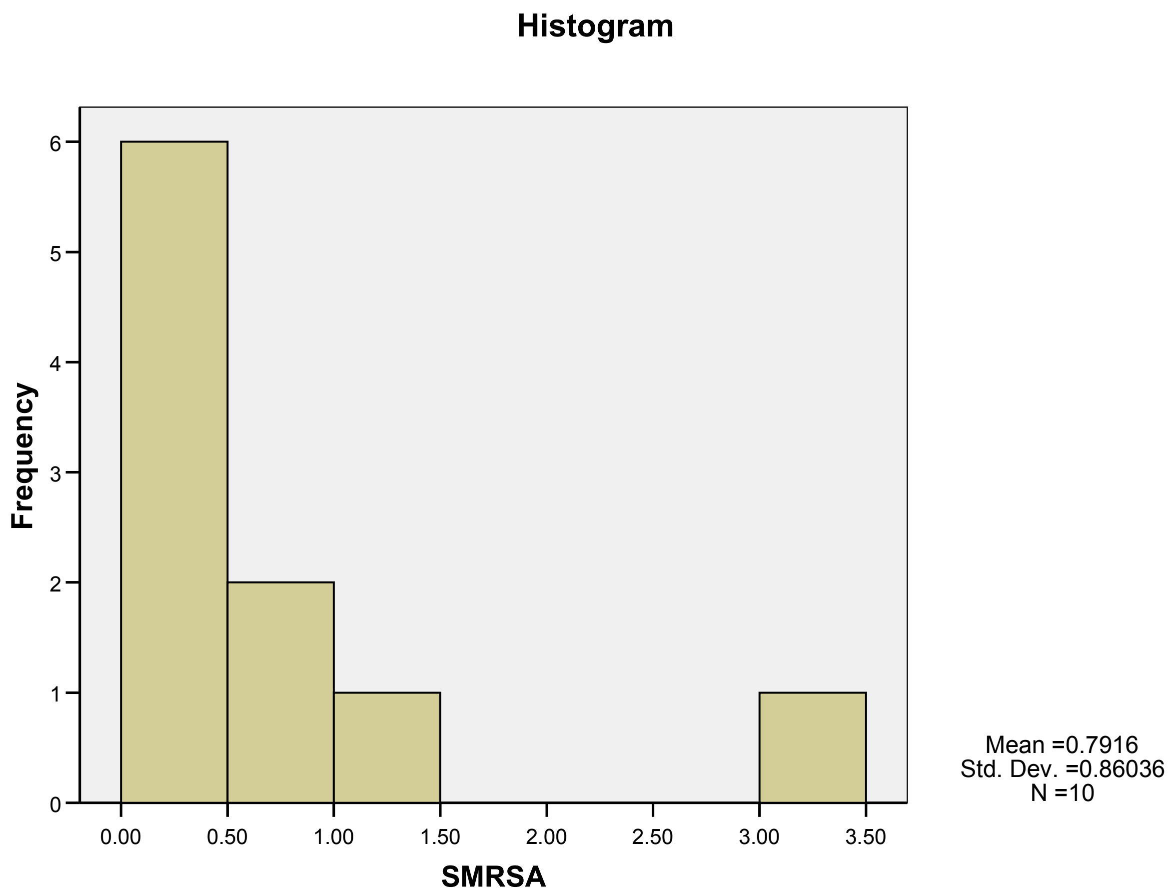
**Distribution of SM-mean IBI scores**



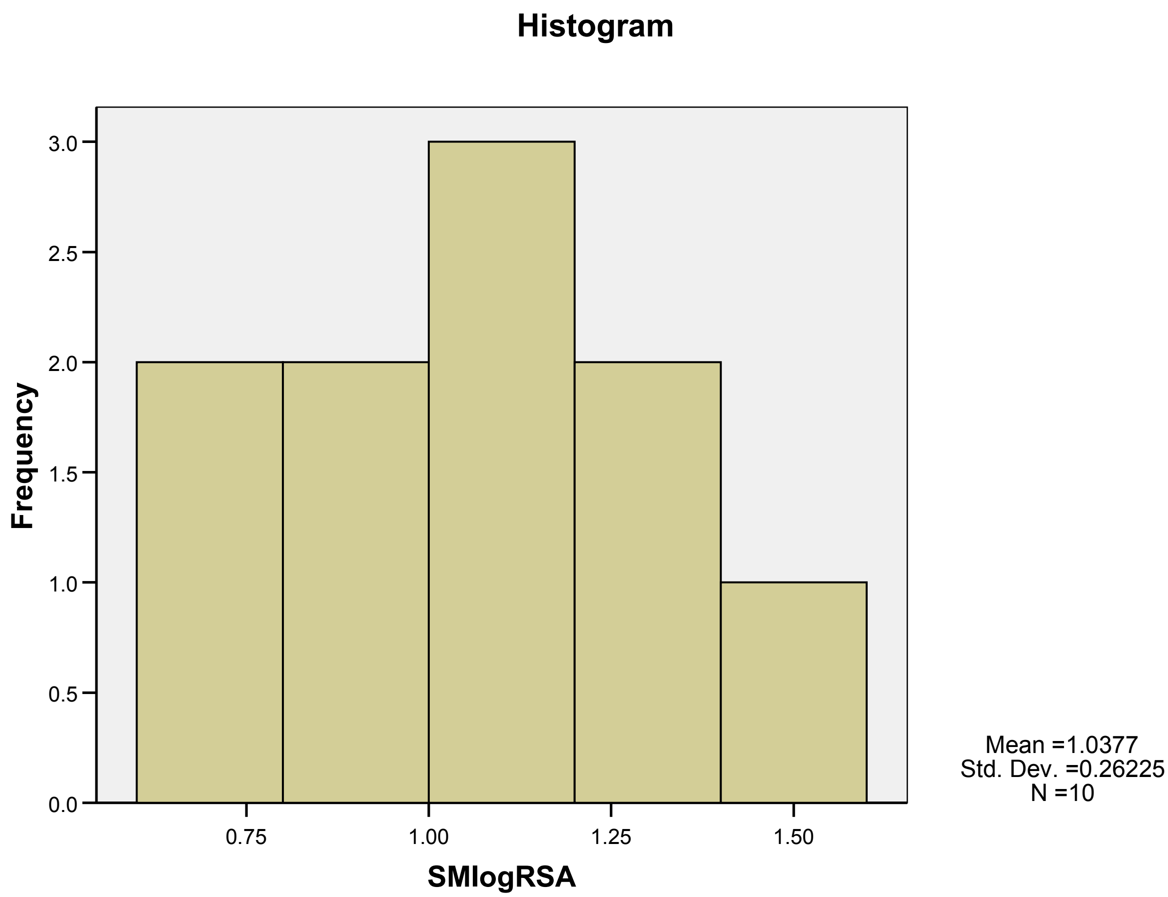
**Distribution of SM-MSD scores**



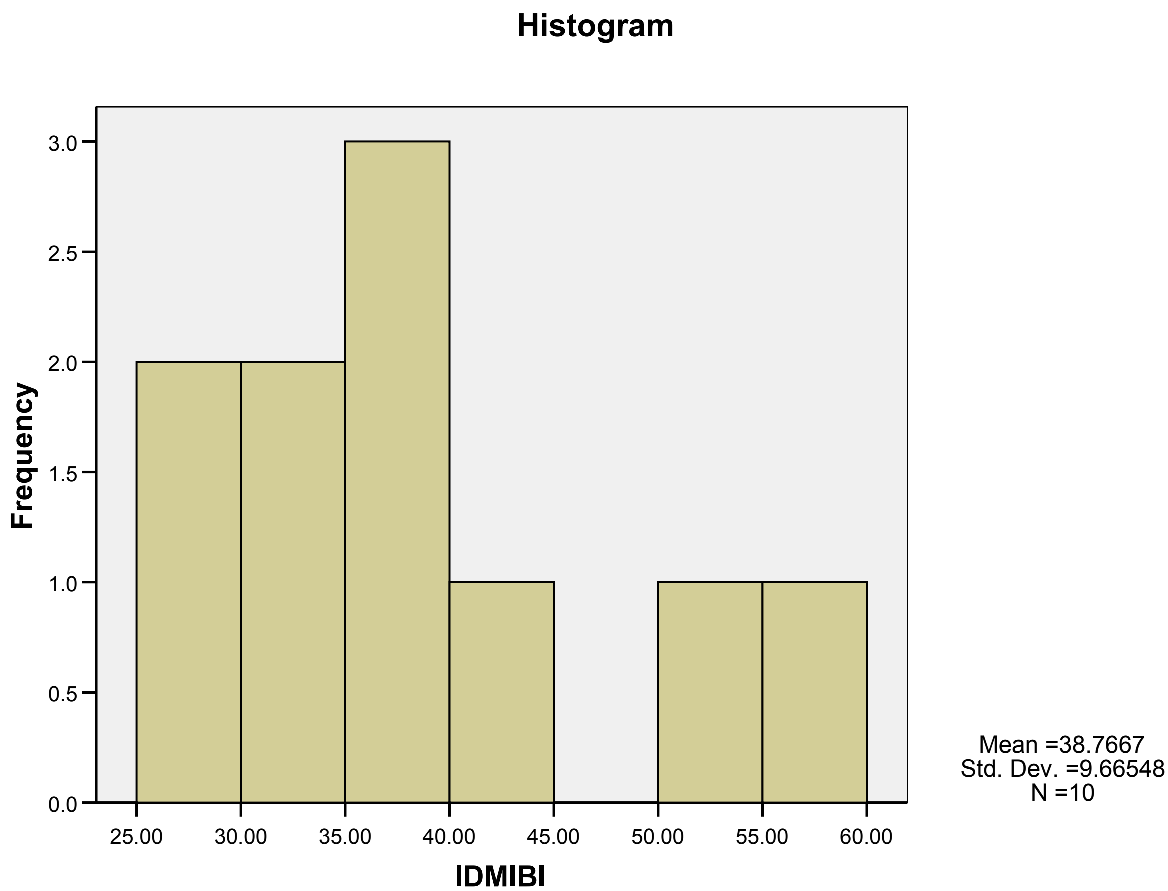
**Distribution of SM-RSA scores**



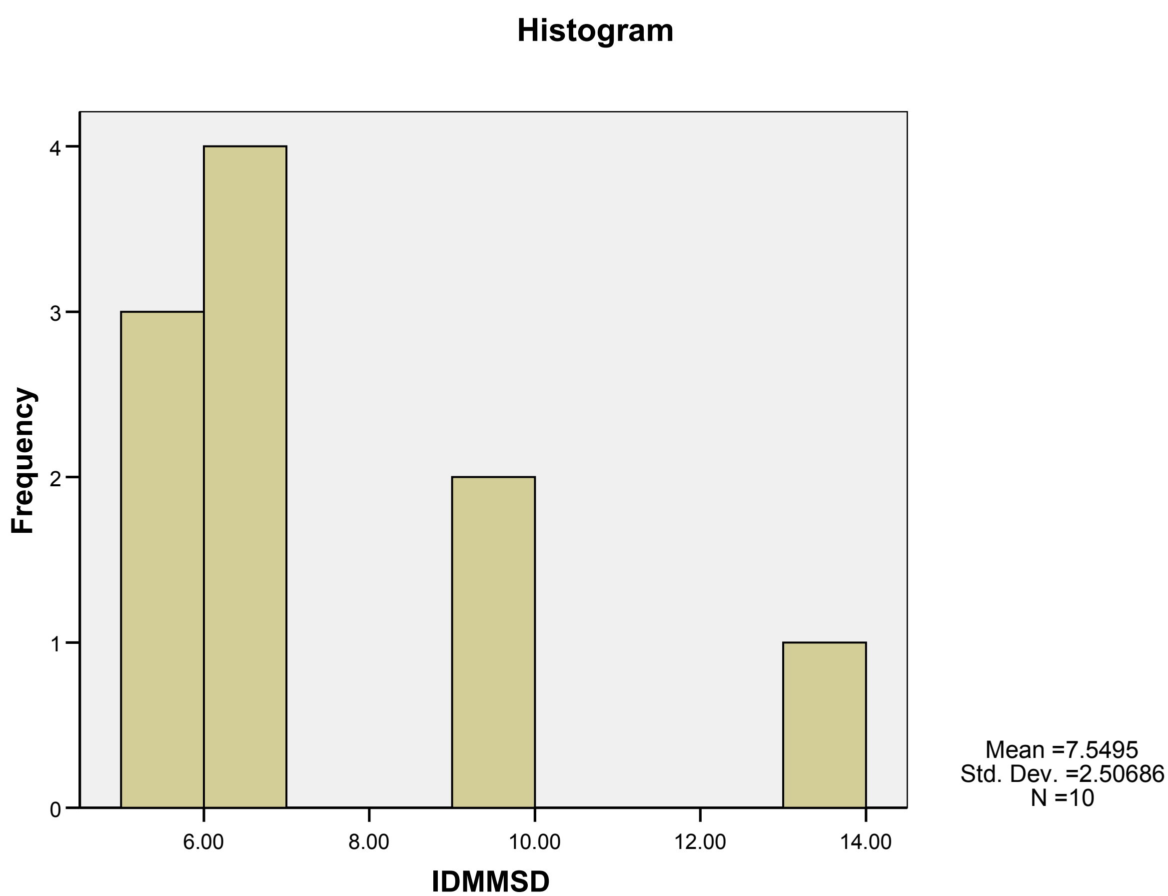
**Distribution of SM-log$_e$ RSA scores**



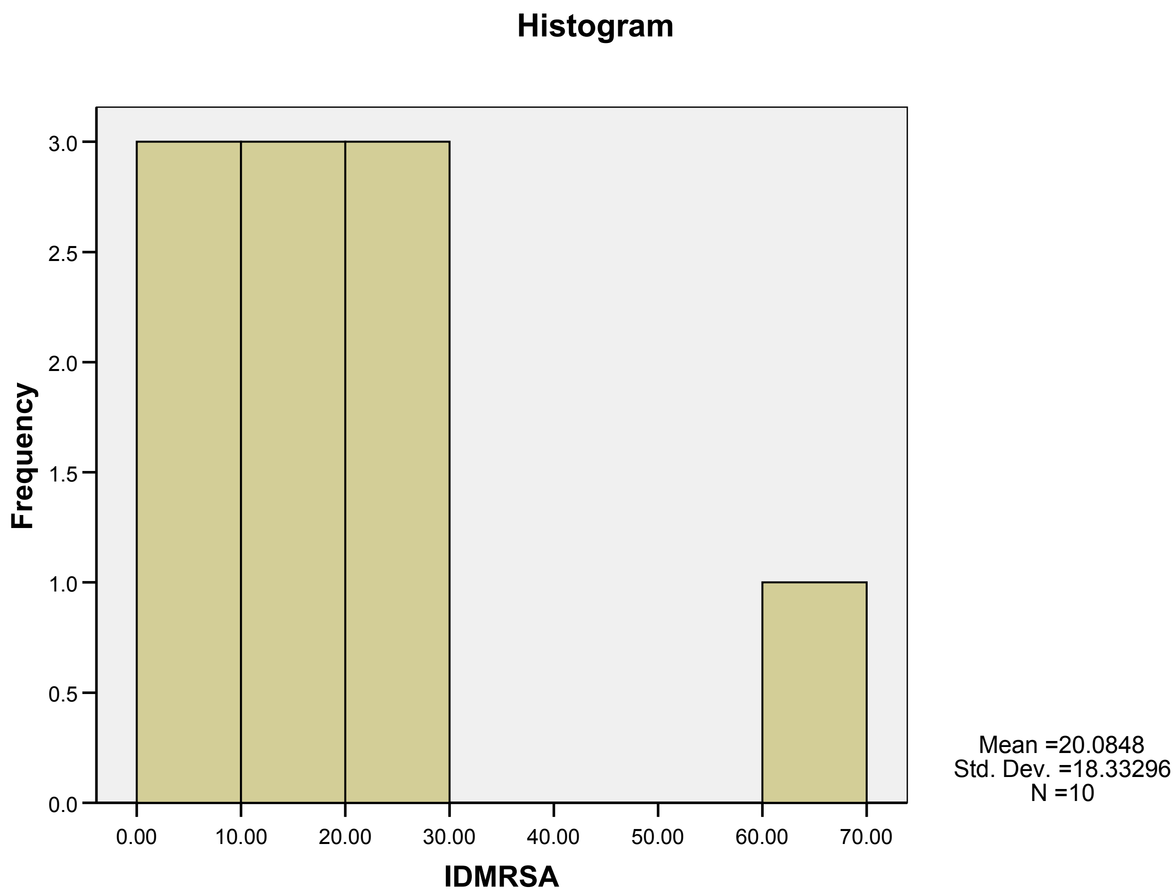
**Distribution of IDM-mean IBI scores**



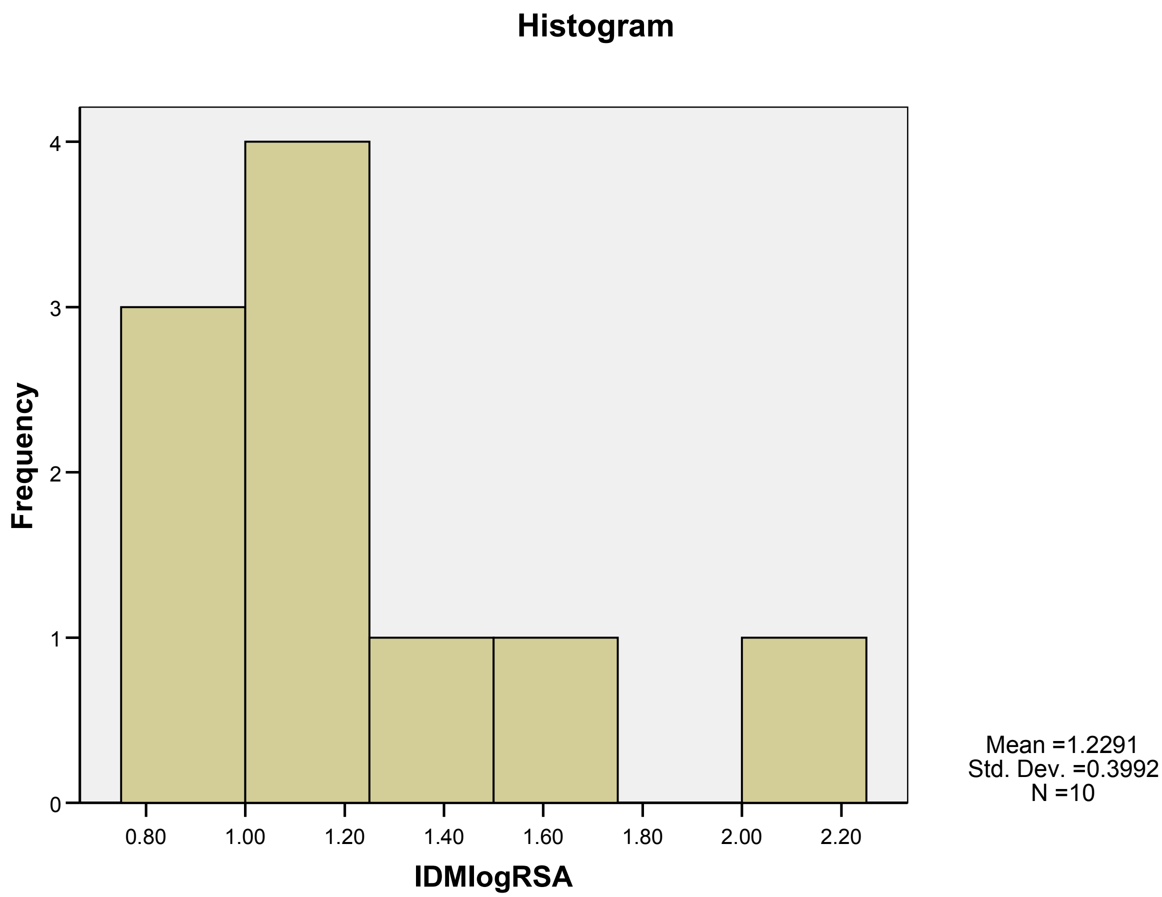
**Distribution of IDM-MSD scores**



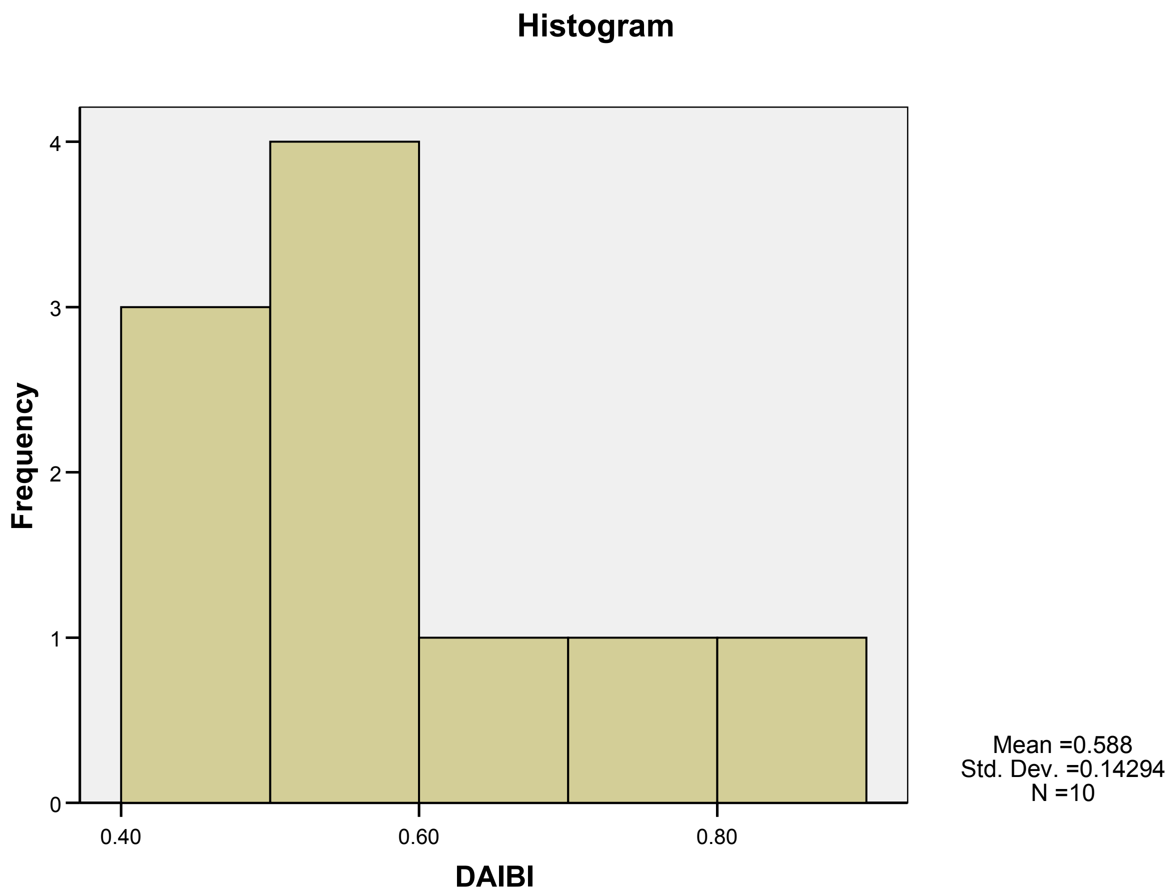
**Distribution of IDM-RSA scores**



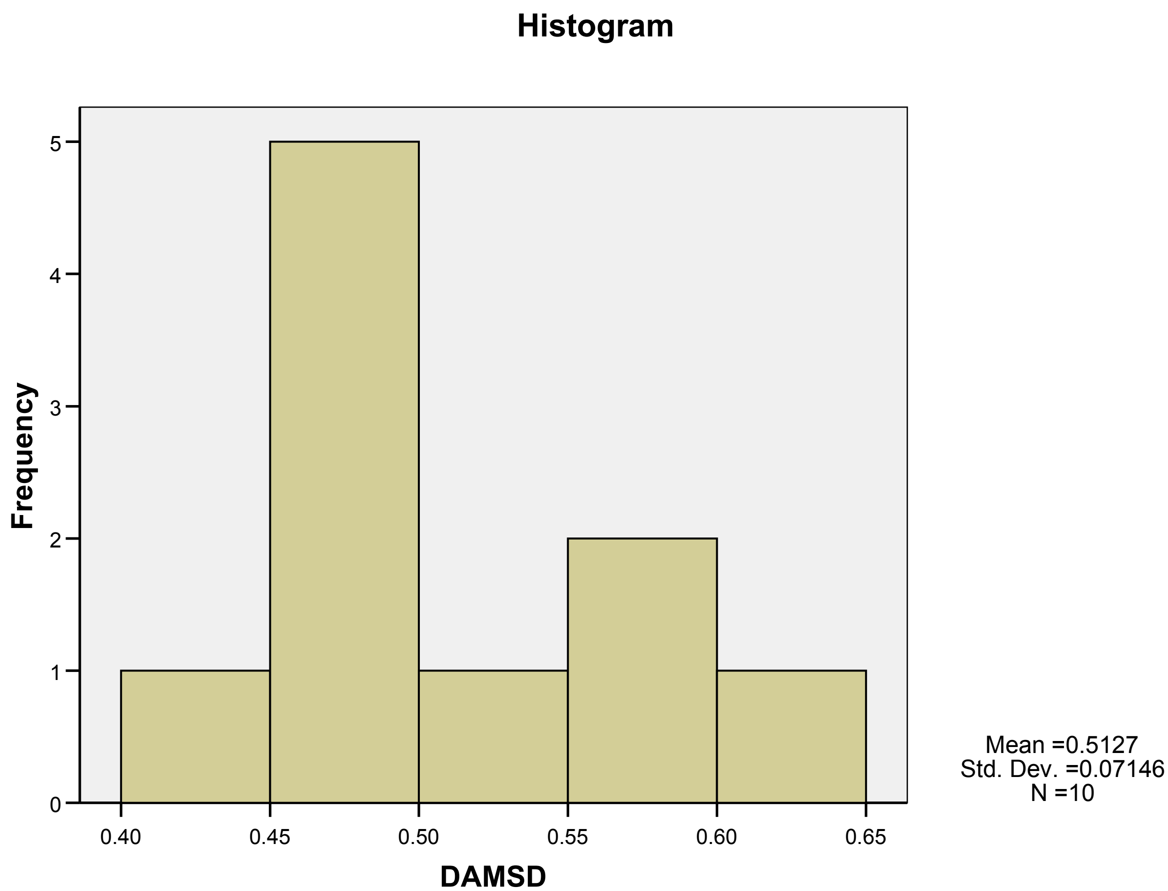
**Distribution of IDM-log$_e$ RSA scores**



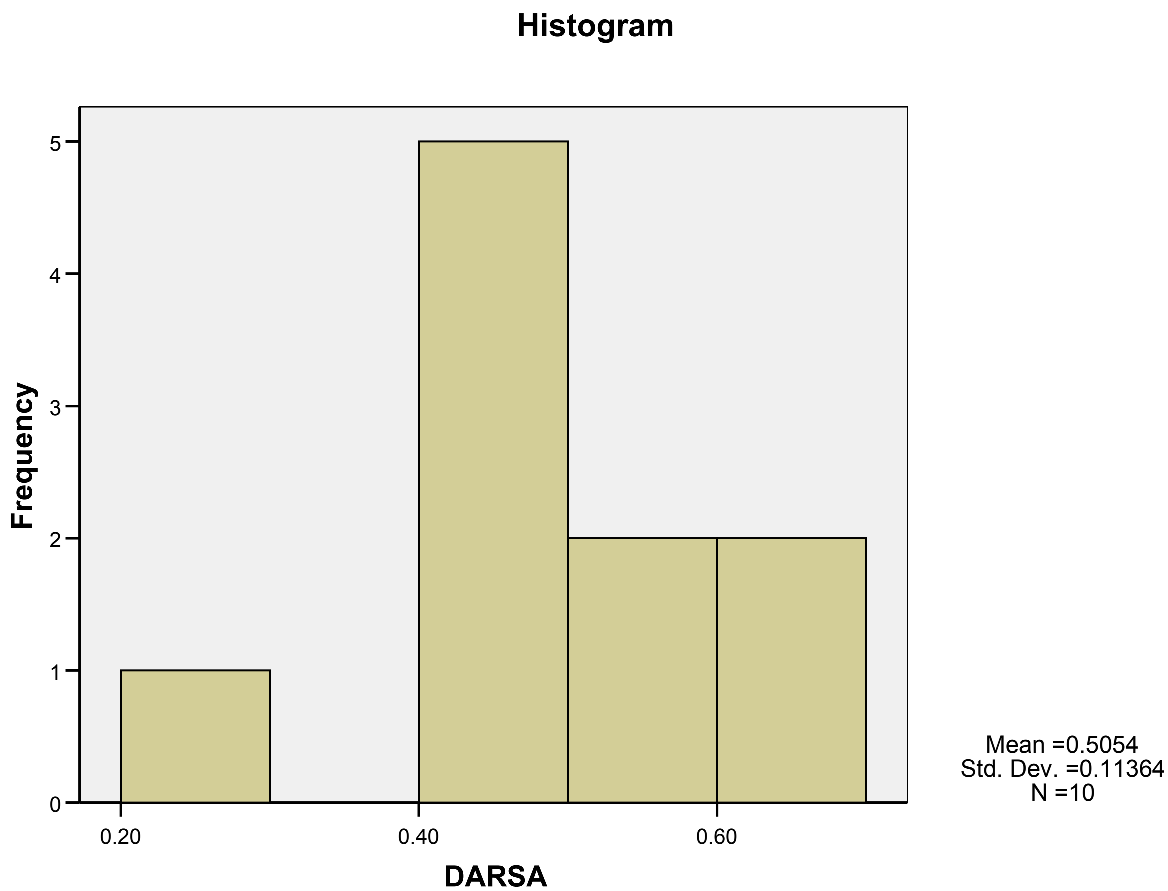
**Distribution of DA-mean IBI scores**



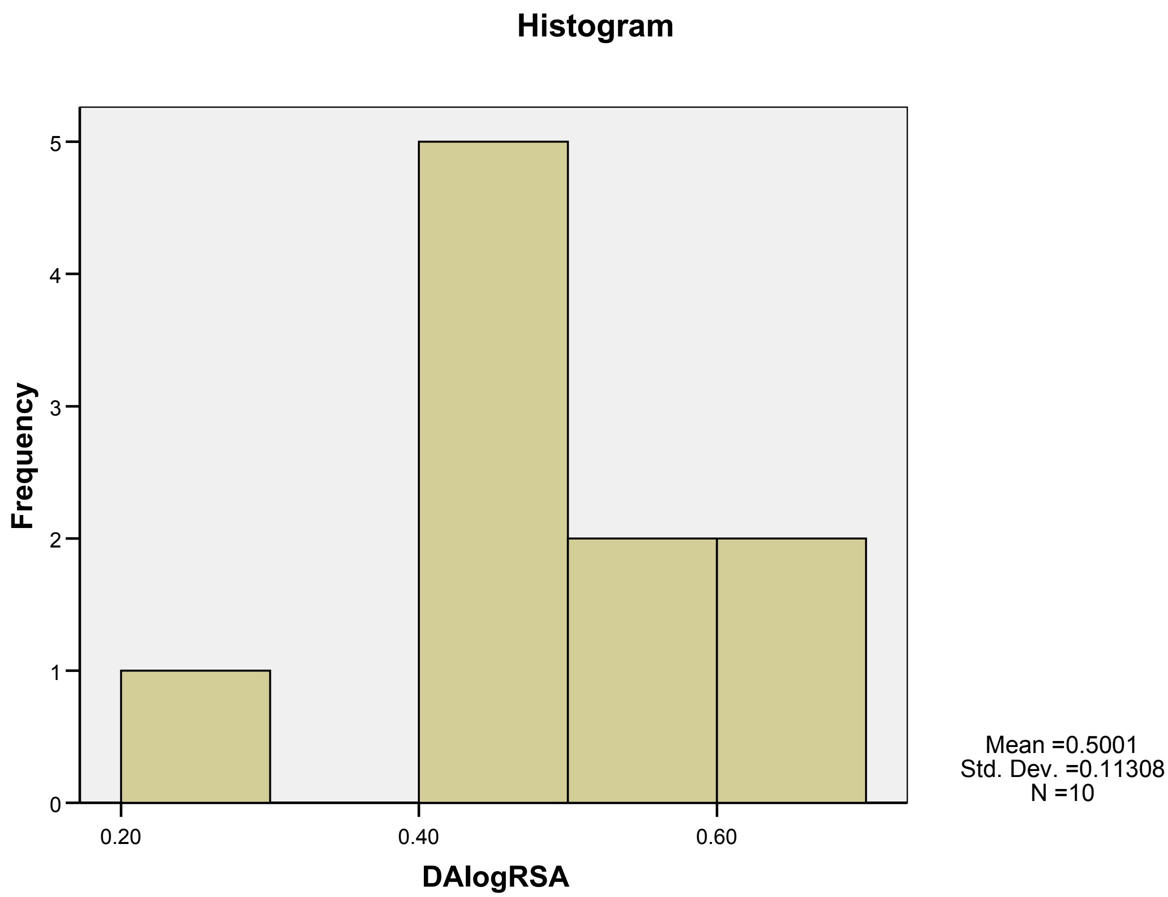
**Distribution of DA-MSD scores**



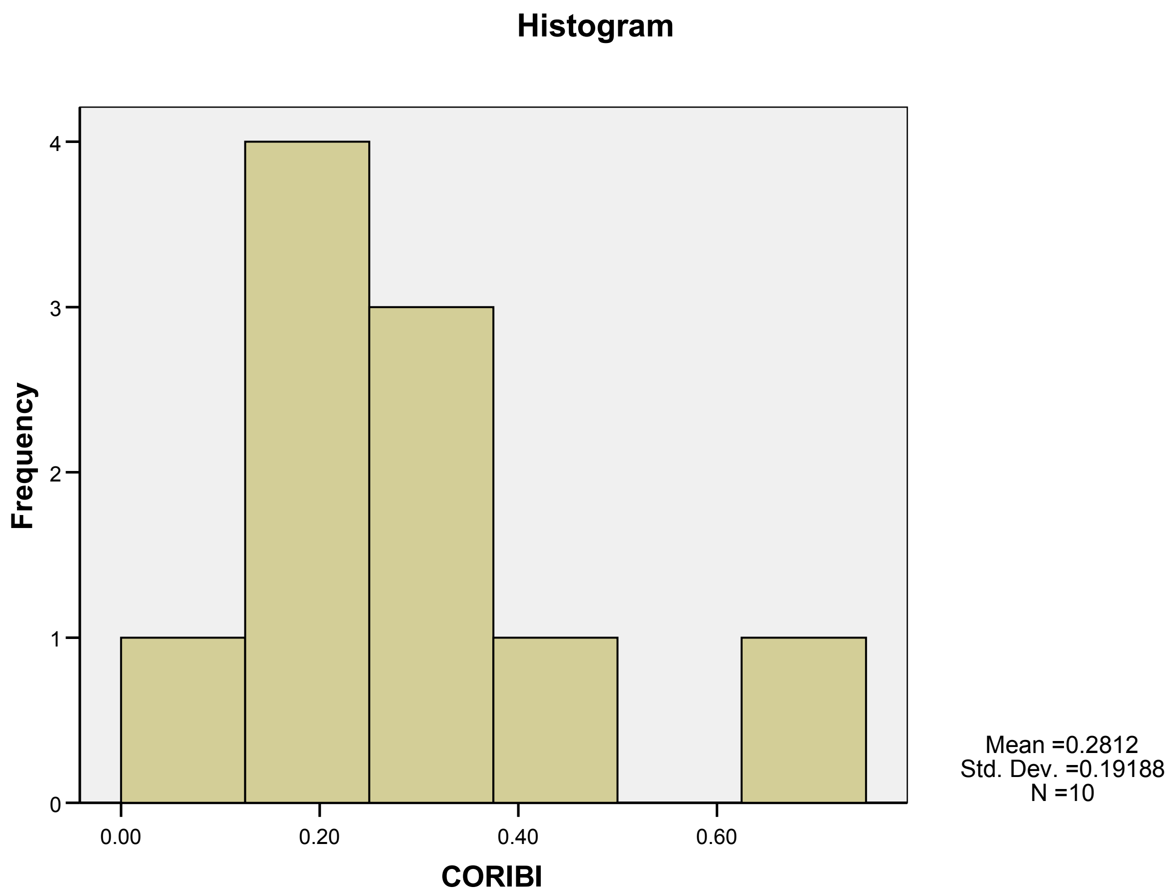
**Distribution of DA-RSA scores**



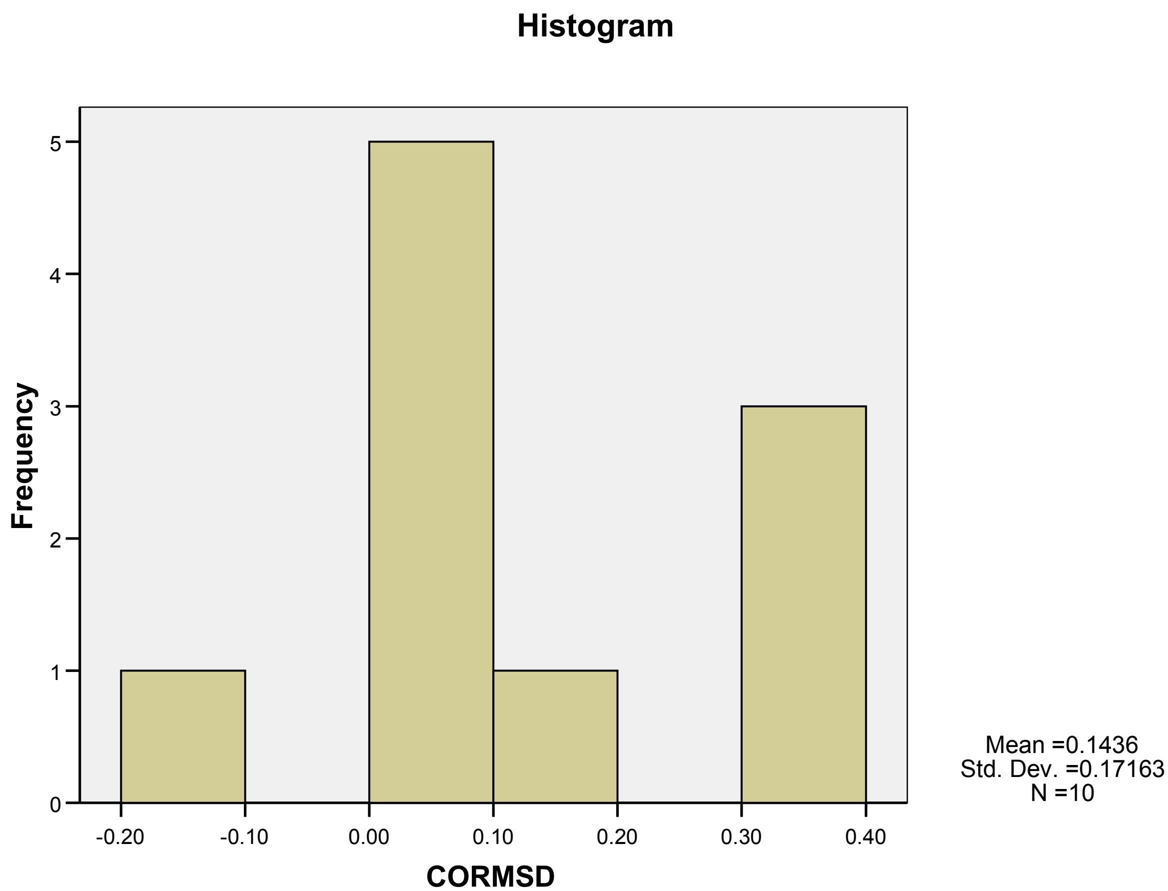
**Distribution of DA-$\log_e$ RSA scores**



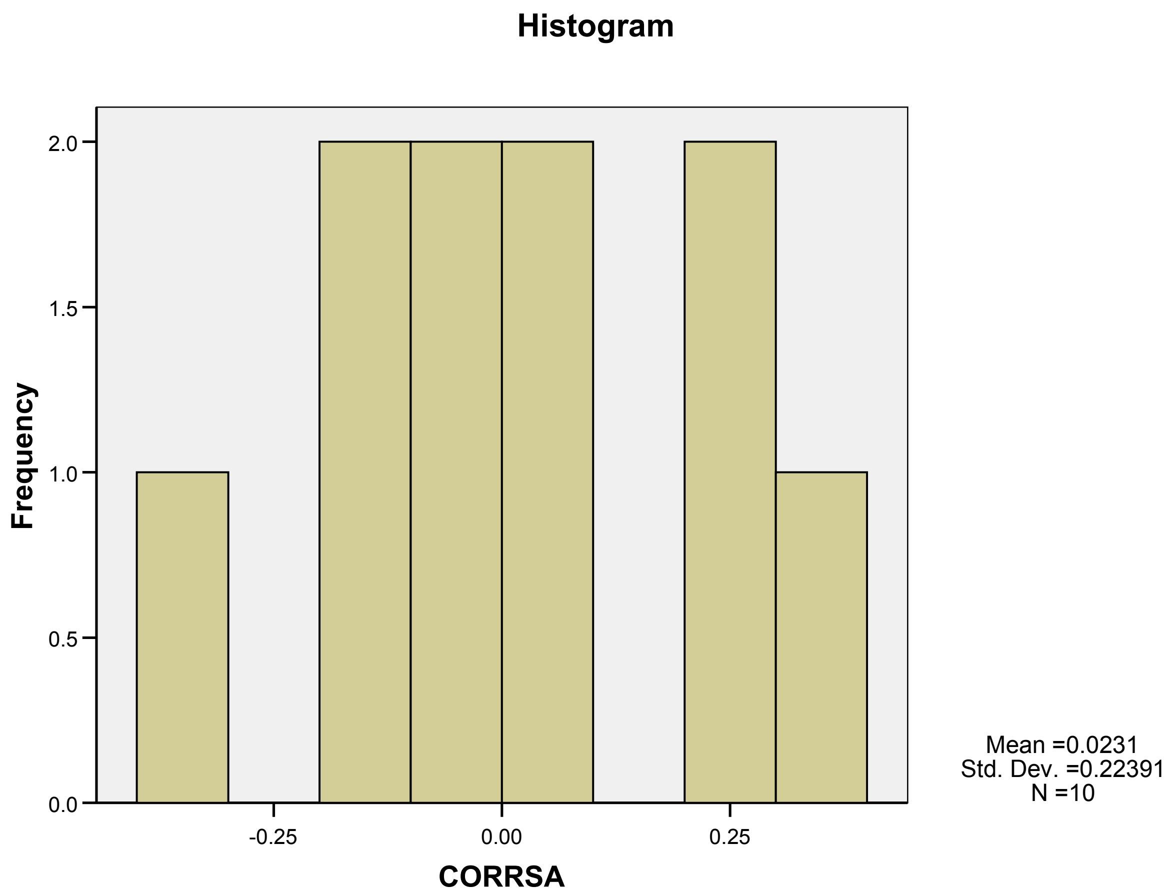
**Distribution of Correlation-mean IBI scores**



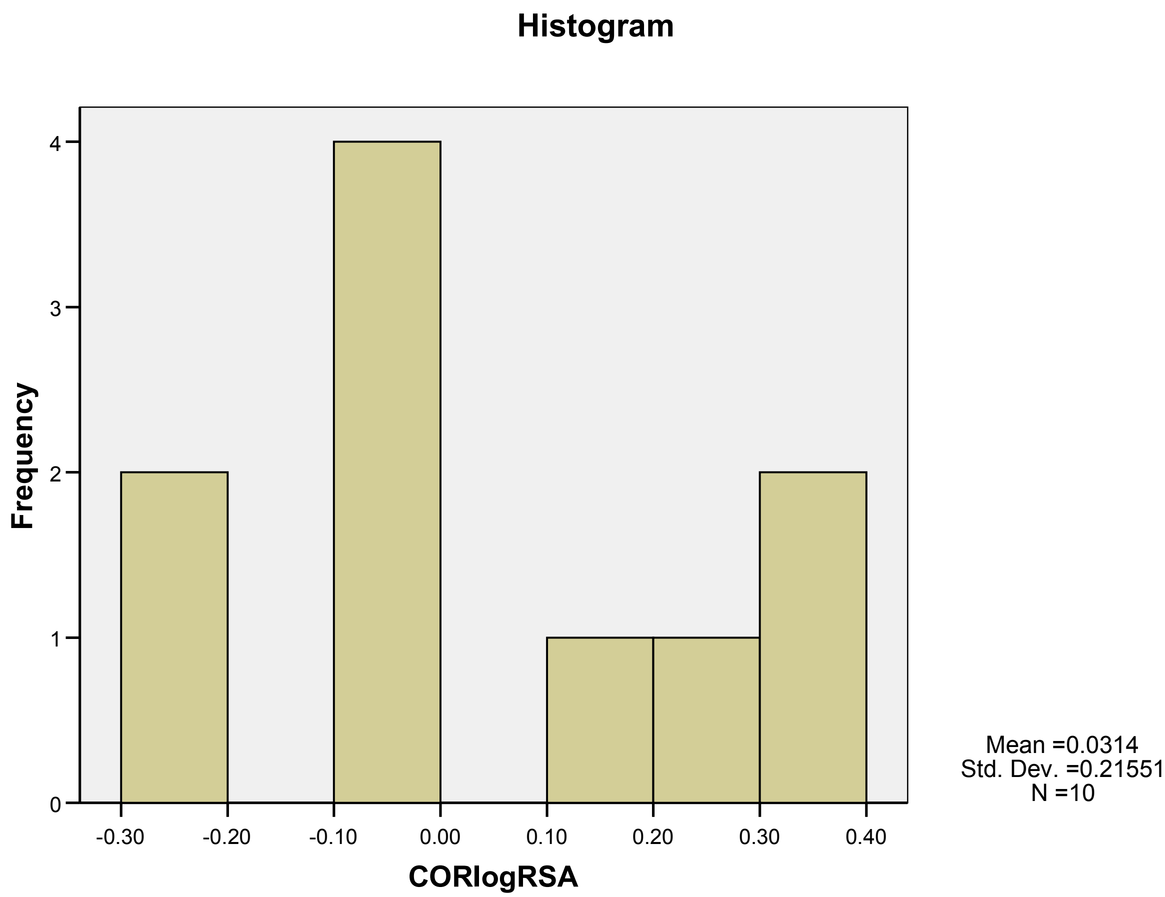
**Distribution of Correlation-MSD scores**



**Distribution of Correlation-RSA scores**



**Distribution of Correlation-log$_e$ RSA scores**



## REFERENCES


Allen, M. T., Matthews, K. A., & Kenyon, K. L. (2000). The relationships of resting baroreflex sensitivity, heart rate variability and measures of impulse control in children and adolescents. *International Journal of Psychophysiology, 37*(2), 185-194.

Andrews, D. (2005). Distributed simulation training for teams. In N. Stanton, A. Hedge, K. Brookhuis, E. Salas, & H. Hendrick (Eds.), *Handbook of Human Factors and Ergonomics Methods,* 45-1 – 45-6. Boca Raton, FL: CRC Press.

Backs, R.W. & Boucsein, W. (Eds.) (2000). Engineering Psychophysiology. Lawrence Erlbaum Associates, Mahwah, NJ.

Beng-Chong, L. & Klein, K. J. (2006). Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy. *Journal of Organizational Behavior, 27,* 403-418.

Berntson, G.G., Bigger, J.T., Eckberg, D.L., Grossman, P., Kaufmann, P.G., Malik, M., Nagaraja, H.N., Porges, S.W., Saul, J.P., Stone, P.H., & Van Der Molen, M.W. (1997). Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology, 34,* 623-648.

Cacioppo, J.T. & Petty, R.E. (1983). Social Psychophysiology: A Sourcebook. Guildford Press, New York.

Carpenter, T. L. (2006). *The effects of game-based virtual environment training on novice team performance of a building clearing task*. Unpublished masters thesis, Clemson University, Clemson, South Carolina, USA.

Chance, P. (1989, May). Great experiments in team chemistry. *Across the Board.* 18-25.

DiMascio, A., Boyd, R.W., Greenblatt, M., & Solomon, H.C. (1955). The psychiatric interview: A sociophysiological study. *Diseases of the Nervous System, 16,* 4-9.

Grossman, P. (1983). Respiration, stress, and cardiovascular function. *Psychophysiology, 20*(3), 284-300.

Hatfield, E., Cacioppo, J.T., Rapson, R.L. (1994). Emotional Contagion. Cambridge University Press.

Hayano, J., Sakakibara, Y., Yamada, A., Mukai, S., Fujinami, T., Yokoyama, K., Wtanabe, Y., & Takata, K. (1991). Accuracy of assessment of cardiac vagal tone by heart rate variability in normal subjects. *American Journal of Cardiology, 67,* 199-204.

Henning R.A., Boucsein W., & Gil M.C. (2001) Social-Physiological compliance as a determinant of team performance. *International Journal of Psychophysiology, 40*, 221-232.

Henning, R. A., Ferris, J. K., & Armstead, A. G. (2006). *Longitudinal study of social psychophysiological compliance in a four-person research team.* Presentation given at the 16th World Congress of the International Ergonomics Association, Maastricht, Netherlands.

Henning, R. A. & Korbelak, K. (2005). Social-psychophysiological compliance as a predictor of team performance. Psychologia, 48(2), 84-92.

Hoover, A., & Muth, E. (2004). A real-time index of vagal activity. *International Journal for Human-Computer Interaction, 17,* 197-209.

Kao, H.S. & Smith, K.U. (1971). Social feedback: determination of social learning. *J. Nervous Mental Dis., 152*, 289-297.

Kaplan, H.B., Burch, N.R., & Bloom, S.W. (1964). Physiological covariation and sociometric relationships in small peer groups. In P.H. Leiderman & D. Shapiro (Eds.), *Psychobiological approaches to social behavior,* Stanford, CA: Stanford University Press.

Kimball, J.W. (2005). *Kimball's Biology Pages.* Retrieved March 15, 2207. Website: http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/P/PNS.html

Levenson, R.W. & Gottman, J.M. (1983). Marital Interaction: physiological linkage and affective exchange. *J. Pers. Social Psychol., 45(3)*, 587-597.

Lucius, R. H., & Kuhnert, K. W. (1997). Using sociometry to predict team performance in the workplace. *Journal of Psychology, 131*(1), 21-32.

Malmo, R.B., Boag, T.J., & Smith, A.A. (1957). Phsyiological study of personal interaction. *Psychosomatic Medicine, 19,* 105-119.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Cannon-Bowers, J. A., & Salas, E. (2005). Scaling the quality of teammates' mental models: Equifinality and normative comparisons. *Journal of Organizational Behavior, 26,* 37-56.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85,* 273-283.

McCraty, R., Atkinson, M., & Tomasino, D. (2001). Science of the heart: Research overview and summaries. Retrieved March 01, 2007 from Institute of HeartMath, HeartMath Research Ceneter Website: http://heartmath.org/research/science-of-the-heart/soh_13.html

Mohammed, S., & Dumville, B. C. (2001). Team mental models in a team knowledge framework: Expanding theory and measurement across disciplinary boundaries. *Journal of Organizational Behavior, 22,* 89-106.

Norman, D. A. (1983). Some observations on mental models. In D. Gentner, & A. L. Stevens (Eds.), *Mental models* (pp. 7-14). Hillsdale, NJ: Erlbaum.

Porges, S. W. (1995). Cardiac vagal tone: A physiological index of stress. *Neuroscience and Biobehavioral Reviews, 19*(2), 225-233.

Reich, R. B. (1987). Entrepreneurship reconsidered: The team as hero. *Harvard Business Review, 65*(3), 77-83.

Salas, E., & Cannon-Bowers, J. A. (1997). Methods, tools, and strategies for team training. In M. A. Quinones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Applications of psychological research* (pp. 249-279). Washington, DC: American Psychological Association.

Salas, E., Milham, L. M., & Bowers, C. A. (2003). Training evaluation in the military: misconceptions, opportunities, and challenges. *Military Psychology, 15*(1), 3-16.

Salas, E., Sims, D., & Burke, C. (2005). Is there a "big five" in teamwork? *Small Group Research, 36*, 555-559.

Sauter, S. & Smith, K.U. (1971). Social feedback: quantitative division of labor in social interactions. *J. Cybernetics, 1(2)*, 80-93.

Smith, K.U. & Smith, M.F. (1966). Cybernetic principles of learning and educational design. Holt, Rinehart, and Winston, New York.

Smith, T.J. & Smith, K.U. (1987). Feedback control mechanisms of human behavior. In: Salvendy, G. (Ed.) *Handbook of Human Factors.* Wiley, New York, 251-293.

Springer, J., Muller, T, Langner, T., Luczak, H., & Beitz, W. (1990). Stress and strain caused by CAD Work – Results of alaboratory study. In L. Berlinguet & D. Berthelette (Eds.), *Proceedings of the Work with Display Units Conference, Montreal,* (231-238). Amsterdam: Elsevier.

Stout, R. J., Salas, E., & Fowlkes, J. E. (1997). Enhancing teamwork in complex environments through team training. *Group Dynamics: Theory, Research, and Practice. 1*(2), 169-182.

Wientjes, C. J. E., ter Maat, R., & Gaillard, A. W. K. (1994). *Workload and stress at the CRC/MilATCC: A test of the Stressomat.* (Rep. No. TNO-TM 1994 A-56 [Dutch]). Soesterberg, The Netherlands: TNO Human Factors Research Institute.