May 6, 2015

Test scores on timed exams decline over time without a significant increase in physiological stress

Timed test performance and physiological stress

Joan Bunyi, Carolyn Heal, Sunita Nadendla, Dawn Sherman, Thuy Dan Tran, and Jacob Varsos

University of Wisconsin – Madison, Department of Physiology
Lab 603
Group 7

3021 words

Abstract

In today's world, there is a lot of pressure to perform well on exams. Schools and workplaces rely on tests to evaluate individuals. While these institutions intend to test for cognitive function, there are a number of variables affecting these evaluations, such as stress, test anticipation, and other physiological factors. Our study examined the relationship between physiological stress and timed cognitive performance, specifically if the stress of test performance positively feedbacks on an individual's respiratory rate, blood pressure, and heart rate. We predicted that these three physiological factors would increase as testing progressed, and that concurrently test performance would decline. Our results indicate that there was no significant change in heart rate or mean arterial blood pressure (MABP), although we did see a significant increase in respiratory rate following the first test section. Our data may therefore lead us to believe that students' responses to testing situations are highly variable and do not collectively follow a general trend. A trend we did find to be statistically significant is declining performance in latter sections of the test. For test takers, it may be helpful to take note of their own stress patterns when taking exams and to adjust their strategies for maintaining calmness in order to potentially maintain high test performance throughout the evaluation.

Keywords: anticipation, blood pressure, cognitive function, performance, stress, test stress

Introduction

Stress is one of the most common factors that contributes to poor academic test performance (Irwin, 1984). Despite this fact, not much is known about the extended effects of stress throughout an academic exam or throughout a series of exams. On the other hand, the relationship between challenging mental activity and anxiety has been extensively studied in the workplace. In these studies, stress and anxiety have been used interchangeably to describe physiological changes that occur in certain situations. Many studies have shown that in demanding work conditions, employees have increased risk for cardiovascular disease primed by cortisol-mediated elevations in heart rate and blood pressure (Krajnak, 2014; Nyberg, 2013; Collins, 2005).

Stress and anxiety are triggered by a number of external factors which include intellectual evaluation, time limits for completing tasks, dealing with aggression, and responding to mortal danger (Kruk, et al, 2003). One of the ways the body responds to these stressors is through the adrenal gland, which secretes cortisol, epinephrine, and norepinephrine. Among their many functions, these hormones stimulate increases in heart rate, blood pressure (Krajnak, 2014), and respiratory rate (Suess, 2007). Other physiological symptoms of stress include reduction in digestion, dilation of pupils and blood vessels, and heightened blood plasma concentrations of glucose and fatty acids (Raff and Levitzky, 2011). Our study is interested in whether a positive feedback relationship exists between test-related anxiety and test performance in college students. A positive feedback system can be described as one factor (Factor A) that greatly affects the output of a second factor (Factor B). As Factor B increases, it feeds back to further increase the output Factor A, which in turn further increases the output of Factor B. This process continues until a terminating event occurs (Jefferson, 1996), at which point the cycle ceases.

This experiment aimed to study the positive feedback relationship between stress and test performance. In general, there has been a significant amount of research on the relationship between test taking and stress. Previous research has focused on how testing has been influenced by the test taker's environment, emotional state, and school curriculum (Zatz, 1985). However, there has been little data collected on whether a positive feedback relationship exists between significantly high stress and test performance. This relationship is important to understand - because in the typical school setting, various subject exams are often scheduled in the same week or even the same day. Ultimately, if the positive feedback loop we've proposed exists, it may be negatively affecting students' academic performance regardless of their preparation or intellect. Results from this experiment can benefit students and educational institutes by providing a better understanding of the relationship between stress and evaluation, and their subsequent effects on test performance. It may also provide suggestions on what strategies students can use in order to attenuate or circumvent the effects of a positive feedback between stress and poor test performance.

Our study presented students with a test divided into three four-minute sections consisting of questions taken from a standardized test. We chose questions from The American College Test (ACT) because it is already a well established national exam with a standardized bank of questions (Hidden curriculum, 2014). Additionally, standardized tests are typically the most common format used in schools throughout the United States. Furthermore, all study participants are currently enrolled at the University of Wisconsin-Madison, which requires a score either from the ACT or the related Scholastic Assessment Test (SAT) prior to admission; all study participants are therefore familiar with the type of material found on the ACT and are capable of taking the test.

We hypothesized that, due to a positive feedback loop between stress and cognitive performance, participants' heart rate, mean arterial blood pressure (MABP), and respiratory rate will increase as the participant progresses from the first test section to the last, while test scores simultaneously decline. We also hypothesized that heart rate, MABP, and respiratory rate will be significantly higher after the last section of the test than the baseline measurements taken before the first section of the test begins.

Materials

- Automatic blood pressure cuff (Omron Healthcare, with ComFit cuff)
- Pulse oximeter (Nonin Medical Inc.)
- Respiratory effort transducer (Biopac Systems, SS5LB)
- Biopac data analysis program
- American College Test (ACT) standardized test math and reading questions
- Proctor script
- Physiology lab consent forms

Methods

Test Development

A three-section exam was developed, consisting of randomly selected mathematical questions and reading passages from retired ACT exams. Each test section consisted of 5 mathematical questions and 5 questions concerning the grammar and comprehension of a short reading passage. The 15 mathematical questions selected for the entire test were randomly distributed into Test 1, Test 2, or Test 3 in an effort to prevent any one section from being more challenging than the others. One reading passage with 15 questions was selected for the entire exam and the first five questions were placed into Test 1; questions 6-10 were placed into Test 2; and questions 11-15 were placed into Test 3. Pilot studies were conducted to ensure that each section was long enough that it could be completed within the four minutes provided, but only with great difficulty. The section lengths were tailored so that most pilot participants completed approximately 70-80% of the questions by the end of the four minutes provided.

Testing Procedure

At the start of the experiment, participants were asked to sit in a relaxed posture with both feet flat on the floor. They were fitted with an automatic blood pressure cuff (Omron Healthcare, with ComFit cuff), a pulse oximeter (Nonin Medical Inc), and a respiratory effort transducer (Biopac Systems, SS5LB). At this time, baseline measures for blood pressure, pulse, and respiratory rate were measured, before any discussion of cognitive examination. These baseline measures were used as our negative control.

The proctor then explained the layout of the exam. Participants were given three test sections with four minutes to complete each. After each section, participants' blood pressure, pulse, and respiratory rate were immediately measured. The proctor clearly explained that the reading passage applies to the entire test and that the participant need only read the part of the reading passage containing questions for that section.

The subject was given the first test section and told to begin. After four minutes, time was called and the first section was confiscated. Blood pressure, pulse, and respiratory rate were immediately recorded. Following this, the second test section was then provided. After the four minute time limit was called, blood pressure, pulse, and respiratory rate was once again immediately recorded. The third and final section was then given. Then, a final recording of blood pressure, pulse, and respiratory rate was taken. Following the exam, each test section was graded and the participant had the option to obtain their test score, but was not required to do so.

All systolic and diastolic blood pressure readings collected by the blood pressure cuffs were converted into mean arterial blood pressures (MABP). The test scores for each section and the physiological measurements taken from all participants were analyzed using Student's paired T-test. Each physiological measure was compared with mean results from the baseline to the first section, from the first section to the second section, and from the second section to the third section. This was done in order to determine whether or not the changes in these physiological measures had statistical significance as the participants moved from section to section.

Controls

Negative: The baseline measurements taken from study participants before the beginning of testing was used as our negative control.

Positive: To confirm that our blood pressure cuff, pulse-oximeter, and respiratory effort transducer were operating correctly, we measured blood pressure, pulse, and respiratory rates in a subject before and after a short burst of physical activity (50 jumping jacks). As expected, our equipment registered increases in blood pressure, heart rate, and respiratory rate immediately following the completion of 50 jumping jacks.

Results

The three physiological measures were collected at baseline and after each of the three sections of the test session. This data was then gathered in a table very similar to Table 1.

Table 2 displays the results of the paired T-tests evaluated for each physiological measure, as well as the final scores of the participants. MABP showed no significant change between any sections of the test, although a slight upward trend can be seen in between the mean arterial blood pressure in Test 1 to Test 2 (p-value = 0. 2429). Overall however, these results indicate that there was no significant change in blood pressure as participants progressed through each test.

The paired T-tests for heart rate also do not show any statistical significance in the data. This indicates that the mean heart rate for the participants did not change significantly throughout the course of the tests. The mean respiratory rate of participants shows statistically significant change between the baseline levels and Test 1, as well as between Test 2 and Test 3 (p-values = . 0133 and . 0217, respectively). The mean difference between baseline and Test 1 indicates that the participants had a significant increase in respiratory rate during this period. However, no statistical difference is seen between Sections 1 and 2, indicating that the participants' respiratory rate remained constant during this period. Subsequently, a significant decrease in respiratory rate is observed after Test 3.

Figures 1, 2, and 3 display boxplots for the MABP, heart rate, and respiratory rate, respectively, for all three sections of the test. Figure 1 shows that the mean MABP for participants does not vary significantly from baseline to the first section of the test, nor from any one section of the test to the next.

Figure 2 shows even smaller differences in the mean heart rate across participants at any point during the testing session than seen in MABP. The upper and lower bounds of the boxplots of each section also indicate very low variability in the heart of the participants in comparison to the other measures used throughout this experiment. The outliers seen in Figure 2 mostly come from a single participant with a higher resting heart rate than the average person, rather than multiple individuals, limiting their effect at skewing results.

Figure 3 shows that the mean respiratory rate varied more noticeably than either the heart rate or the MABP. This figure shows that mean respiratory rate rises after Test 1 from baseline, remains constant at this level through Test 2, then significant drops after Test 3.

The mean scores of participants can be seen in Figure 4. The mean scores clearly drop as one moves from Test 1 to 3, matching the expected results that were enumerated in the hypothesis of this study. Although there a few outliers that can be seen in Figure 4, there are too few for these data points to significantly skew the data.

Discussion

Although test scores significantly declined from the first section to the third, our data did not clearly reflect the steady increase in physiological stress that we expected. We found no significant change in heart rate or MABP, although we did see a significant increase in respiratory rate following the first test section. Our data shows there is no clear positive feedback loop in which declines in test performance and elevations in physiological stress amplify one another. Rather, the transient spike in respiratory rate may indicate that when students are taking timed exams, any stress response may peak early and then decline as the student "settles" into the testing situation.

Beyond that, our data may only indicate that students' responses to testing situations are highly variable and do not collectively follow a general trend. Many studies have been conducted on the theory "warriors versus worriers" in dealing with stressful situations, including academic testing environments. "Warriors" utilize stress in order to do especially well in stressful situations, whereas with "worriers," the anxiety causes them to perform significantly worse (Larner, 2008; Henricsson, 2006). Further evidence suggests that the dichotomy may be influenced genetically, especially at the catechol-O-methyltransferase (COMT) gene. Those with COMT alleles that have valine at the 158th codon produce lower amounts of dopamine in the prefrontal cortex of their brain than those with methionine at the same codon, and as a result may be better at coping with stressful situations (Buckert et al, 2012; Stein et al, 2006). It may be advisable for students to take note of their own stress patterns when taking exams and to adjust their strategies for maintaining calm as needed.

Since our data does show a significant decline in test performance from the first section to the last, it offers some recommendations for students. If the format of an exam is provided, it may be advisable to focus studying especially on the topics known to be covered in the latter sections of the exam. This is especially prudent for standardized exams, whose format is relatively constant. For example, students taking the ACT may prefer to spend some extra time studying for Science section, as it is the last multiple choice section on the exam. Likewise, a student taking the SAT may wish focus especially the grammar covered in the Writing section, and a student taking the MCAT may devote more time to the final section on psychology and sociology.

It is important to comment on the taking of standardized tests and the ACT. All participants were college juniors or seniors, meaning the last time one took the ACT could have been three years or more prior to the study. Younger participants may have been more familiar with the ACT, making our test sections less stressful and easier to take than it was for older participants. Additionally, we did not request the participants' previous ACT scores or collect information about their experience with standardized tests. Some feel standardized tests are not an accurate assessment of one's intelligence and that succeeding on them more involves strategy than actual knowledge. Having only used standardized math and reading sections, we may have limited participants' performance if some prefer the science or English ACT sections.

The determination of respiratory rate had a few confounding factors that needed to be resolved throughout this experiment. Because the subjects had single or multiple layers of clothing, it was possible the reading of chest activity was hampered throughout the experiment.

For several participants, only a minimal level of activity was observed via the respiratory effort transducer, making the distinction of individual breaths more difficult. This problem can be observed in figure 5, which explains the process of determining respiratory rate. Utilizing this process, some respiratory rates were determined to be somewhat higher than physiologically expected levels, but these rules were held across all participants in order to remain consistent and to avoid further confounding the results. Additionally, at the end of Test 3, many participants took a long breath or held their breath after finishing; this may have resulted in a lower observed respiratory rate than expected.

Besides the respiratory equipment, other inconsistent equipment data collection could be a potential error. We were not always working with the same heart rate monitor, respiration belt and blood pressure cuff for all our participants. Each different piece of equipment may have produced readings somewhat inconsistent with the others. Additionally, there could be errors due to the delay it took for physiological data to be taken after each test of the experiment was completed by the subject in that heart rate, blood pressure, and respiratory rates could have reverted back to their resting physiological baselines. Other limitations include having limited the participant population to only students enrolled in physiology 435. We also have a somewhat small sample size, which could affect the reliability and validity of our results.

In addition, these participants were aware that their performance during the experiment did not affect them personally - that is, there were no high stakes in this testing situation as there would be with a real exam. This could have lead to participants not being fully committed to performing well, thus skewing test results. It is possible that a different pattern of physiological stress and performance exists in real testing situations like the ACT, college exams, or the MCAT, where the perceived stakes are much higher. Future studies may get a more accurate estimate of stress-performance relationships by finding various ways to increase the stakes of the experimental testing situation. For example, many of the students in Physiology 435 are pre-medical students who still need to take the MCAT exam. If the test developed by the experimenters included questions from the MCAT, one may tell participants that their results on the experiment's test may be an indicator of how they'll do on the actual MCAT. This may be an incentive for the participants to try as hard as they can in order to perform well on the experiment's test, resulting in a better indicator of the stress-performance relationship. Another strategy to make the experimental test "higher stakes" is to tell participants that they will be given a monetary reward if they score higher than a certain percent correct, or if they are one of the top scorers in the sample.

To further understand the relationship between stress, physiology, and cognitive performance, we recommend some continued studies. One useful investigation would be to follow the same procedure using other standardized tests such as SAT, GED, MCAT or GRE and comparing the results to this study with the ACT. Using different standardized tests would be useful in understanding whether stress responses remain consistent or vary across different exams. Furthermore, it would be interesting to see what the effects are of adding short breaks between sections or of exercising prior to test taking. Both these changes could potentially relieve stress and allow the participant to have higher test performance. Finally, it would be important to see how stress responses compare between exams that are divided into sections and tests that are given all in one chunk. If sectioned exams result in a significantly greater or lower stress response than an undivided exam, then it may provide recommendation to instructors on how to format their tests.

References

Buckert M, Kudielka BM, Reuter M, Fiebach CJ. (2012). The COMT Val158Met Polymorphism Modulates Working Memory Performance Inder Acute Stress. *Psychoneuroendocrinology*. **37**(11): 1810-1821

Collins SM, Karasek RA, Costas K. (2005). Job Strain and Autonomic Indices of Cardiovascular Disease Risk. *Am J Ind Med*. **48**(3): 182–93

Henricsson L. (2006). Warriors and Worriers: Development, Protective and Exacerbating Factors in Children with Behavioral Problems. A Study Across the First Six Years of School. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences.* Uppsala Universitet

Hidden curriculum (2014). In S. Abbott (Ed.), The Glossary of Education Reform. Retrieved from http://edglossary.org/hidden-curriculum

Irwin, SG. (1984). Stress, Anxiety, and Cognitive Interference: Reactions to Tests. *JPSP*. **46**(4): 929-938.

Jefferson D, Zuckerman B. (1996). Human Population and the Environmental Crisis. *Jones & Bartlett*. p.42

Jones, GM, Jones BD, Hardin B, Chapman L, Yarbrough T, Davis M (1999). The Impact of High-Stakes Testing on Teachers and Students in North Carolina. *The Phi Delta Kappan*. **81**(3): 199-203.

Krajnak KM. (2014). Potential Contribution of Work-Related Psychosocial Stress to the Development of Cardiovascular Disease and Type II Diabetes: A Brief Review. *Environ Health Insights*. **8**(1): 41-45

Kruk MR, Halász J, Wout M, Haller J. (2003). Fast Positive Feedback Between the Adrenocortical Stress Response and a Brain Mechanism Involved in Aggressive Behavior. *US: American Psychological Association*. **118**(5), 1062-1070

Larner, C. (2008). On the Making of Warriors Out of Worriers: The Management of Trait Anxiety in Competitive Sports. *Victoria University Research Depository.*

Nyberg ST, Fransson EI, Heikkilä K, Alfredsson L, Casini A, Clays E, et. al. (2013). Job Strain and Cardiovascular Disease Risk Factors: Meta-analysis of Individual-participant Data from 47, 000 men and women. *PLoS One*. **8**(6): e67323.

Raff H, Levitzky M. Medical Physiology: A Systems Approach. McGraw-Hill, 2011. Print.

Stein DJ, Newman TK, Savitz J, Ramesar R. (2006). Warriors versus Worriers: The Role of COMT Gene Varients. *CNS Spectrums.* **11**(10): 745-748

Suess MW, Alexander AB, Sweeney HW, Marion RJ. (2007). The Effects of Psychological Stress on Respiration: A Preliminary Study of Anxiety and Hyperventilation. *Psychophysiology*. **17**(6): 535-540

Zatz, Sheri; Chassin, Laurie (1985). Cognitions of test-anxious children under naturalistic test-taking conditions. *Journal of Consulting and Clinical Psychology*. **53**(3): 393-401.

Acknowledgements
We would like to acknowledge the guidance and support from Dr. Andrew Lokuta, the physiology 435 PLVs and teaching assistants, our statistician, Nick Zaborek, and all those who participated in our study.

Tables and Figures

Table 1: Mean and standard deviations (s.d.) of MABP*, heart rate, respiratory rate, and test scores (n=26)

	Baseline		
	MABP	Heart Rate	Respiratory Rate
Mean	82.58	76.00	17.34
S.d.	8.90	11.08	3.61
	Test 1		
	MABP	Heart Rate	Respiratory Rate
Mean	81.55	76.77	20.18
S.d.	9.05	14.76	4.84
	Test 2		
	MABP	Heart Rate	Respiratory Rate
Mean	79.78	77.69	20.13
S.d.	8.55	13.49	4.62
	Test 3		
	MABP	Heart Rate	Respiratory Rate
Mean	80.44	78.31	17.57
S.d.	8.57	13.36	4.80
	Scores		
	Test 1	Test 2	Test 3
Mean	6.12	4.96	4.38
S.d.	1.42	1.73	1.65

* Mean Arterial Blood Pressure

Table 1 shows the mean and standard deviation of the data collected throughout the experiment for all participants across all the three test sections. MABP, heart rate, respiratory rate, and test scores were measured.

Table 2. Paired T-tests between physiological measures between each test (n=26)

MABP*	P-value	Mean of Differences:
Initial vs. Test 1	0.5633	1.0256
Test 1 vs. Test 2	0.2429	1.7692
Test vs. Test 3	0.5722	-0.6667
Heart Rate	**P-value**	**Mean of Differences:**
Initial vs. Test 1	0.6524	-0.7692
Test 1 vs. Test 2	0.6134	-0.9231
Test 2 vs. Test 3	0.6869	-0.6154
Respiratory Rate	**P-value**	**Mean of Differences:**
Initial vs. Test 1	0.0133	-2.6923
Test 1 vs. Test 2	0.9522	0.0538
Test 2 vs. Test 3	0.0217	2.5615
Scores	**P-value**	**Mean of Differences:**
Test 1 vs. Test 2	0.0131	0.1154
Test 2 vs. Test 3	0.1848	0.0462
Test 1 vs. Test 3	0.0001	0.1615

*Mean Arterial Blood Pressure

Table 2 shows the results of the Paired T-tests evaluated between the baseline (initial) physiological measures and the first test, as well as between the subsequent tests. Mean arterial blood pressure (MABP), heart rate, and respiratory rate were evaluated for statistical significance. The results of T-tests performed on score (i.e., performance) are also shown in this table.

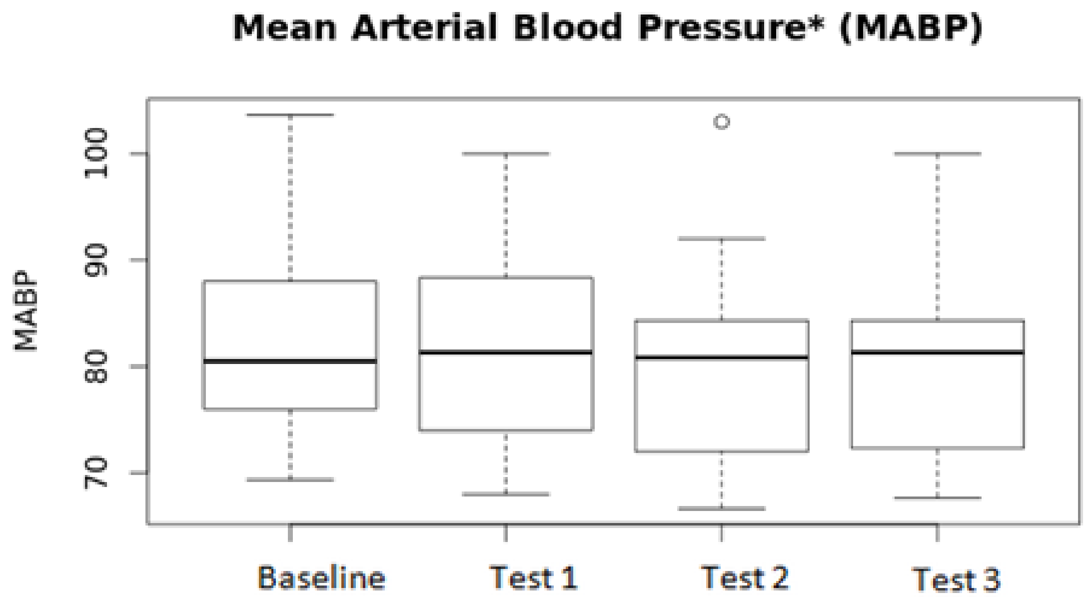

Figure 1: Boxplots of Mean Arterial Blood Pressure (MABP) From Baseline to Test 3 (n=26)

In this figure, the mean MABP for each test can be found within the upper and lower bounds of the box. The black line indicates the median for that subset of MABP data and indicates particular trends in the data similarly to mean. The mean values for baseline, Test 1, Test 2, and Test 3 were 82,58, 81.55, 79.78, and 80.44 respectively. The boxes, and thus the medians, in this boxplot follow the same trend seen in the mean MABP values. The minimum and maximum data points are indicated by the horizontal lines attached to each box by a dotted line, with outliers indicated by circles. The upper bound of the box is 25% greater than the median, while the lower bound of the box is 25% less than the median, outlining the lower and upper quartiles for each data subset. In this particular boxplot, no significant changes in mean MABP are shown.

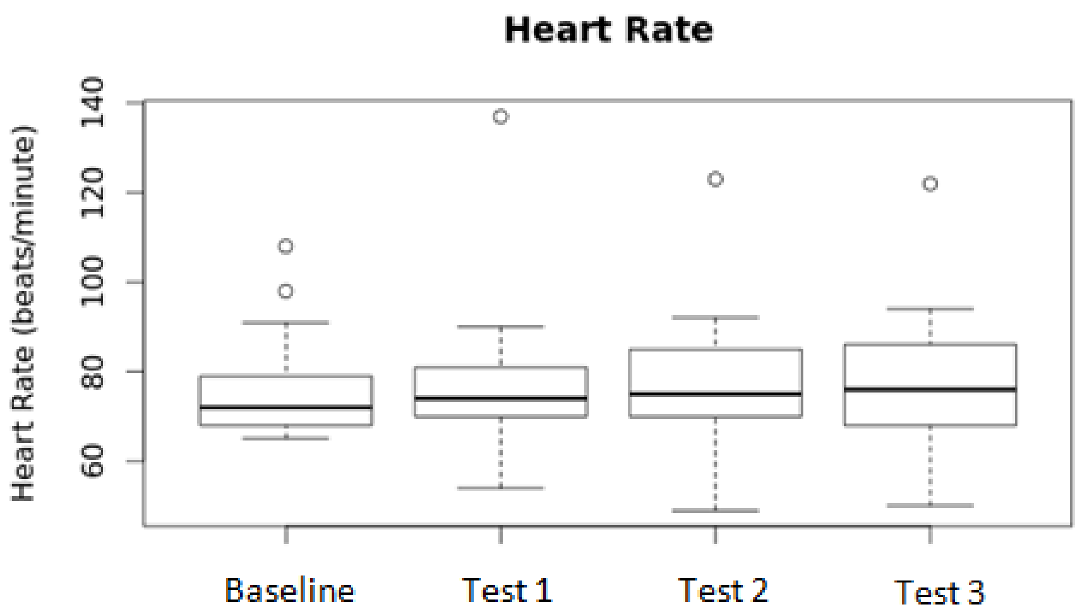

Figure 2: Boxplots of Heart Rate from Baseline to Test 3 (n=26)

The means for baseline, Test 1, Test 2, and Test 3 were 76.00, 76.77, 77.69, and 78.31 respectively. These means are depicted similarly by the medians in these boxplots, and thus their general trends can be seen above. No significant change in mean heart rate was observed as participants progressed through each test. Refer to Figure 1 for a descriptive interpretation of boxplots.

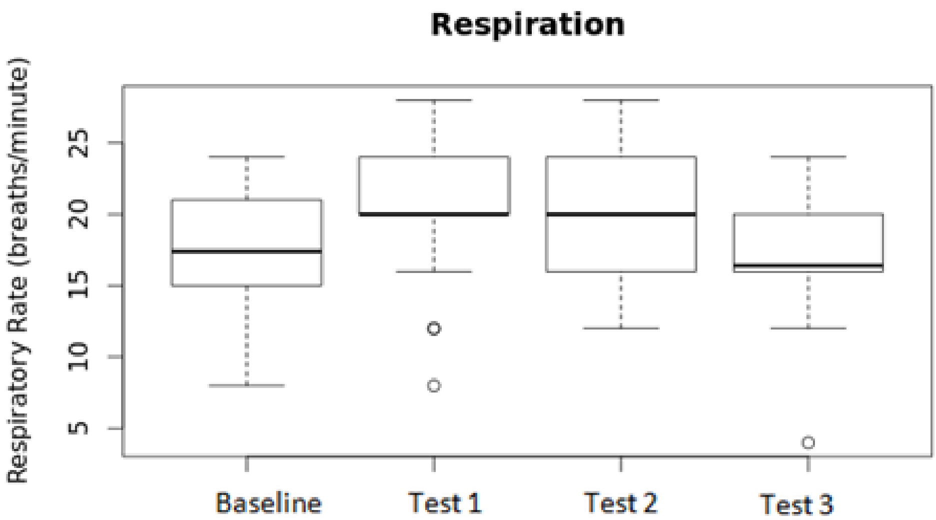

Figure 3: Boxplots of Respiratory Rate from Baseline to Test 3 (n=26)

Figure 3 shows that the mean respiratory rate varied more noticeably than either the heart rate or the MABP. The mean MABP for baseline, Test 1, Test 2, and Test 3 were 17.34, 20.18, 20.13, and 17.57 respectively. This figure shows that mean respiratory rate rises significantly after Test 1 from baseline, remains constant at this level through Test 2, then significant drops after Test 3. Refer to Figure 1 for a descriptive interpretation of boxplots.

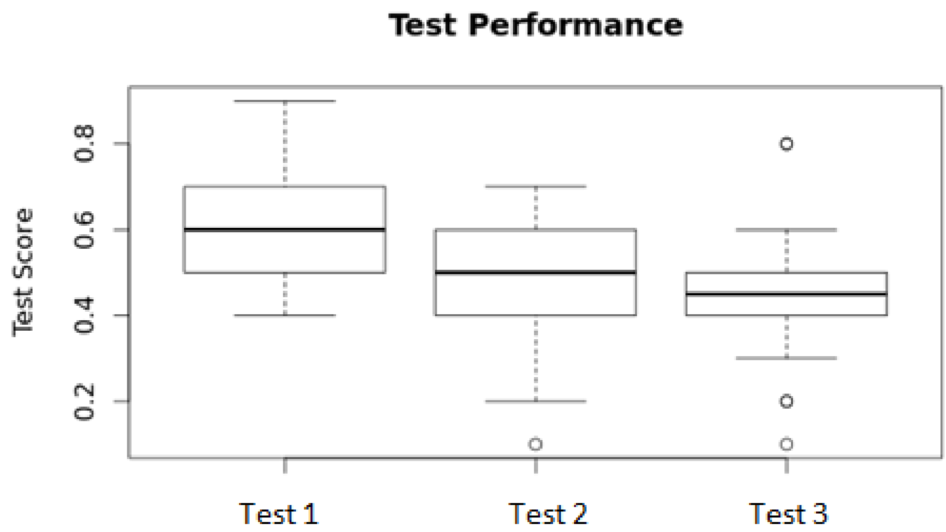

Figure 4: Boxplots of Scores across All Three Tests (n=26)

The mean scores of participants can be seen in Figure 4. The means for Test 1, Test 2, and Test 3 were 6.115, 4.961, and 4.385 respectively. A significant decrease is seen from Test 1 to Test 2 in mean score. Although there a few outliers that can be seen in Figure 4, there are too few for these data points to significantly skew the data. Refer to Figure 1 for a descriptive interpretation of boxplots.

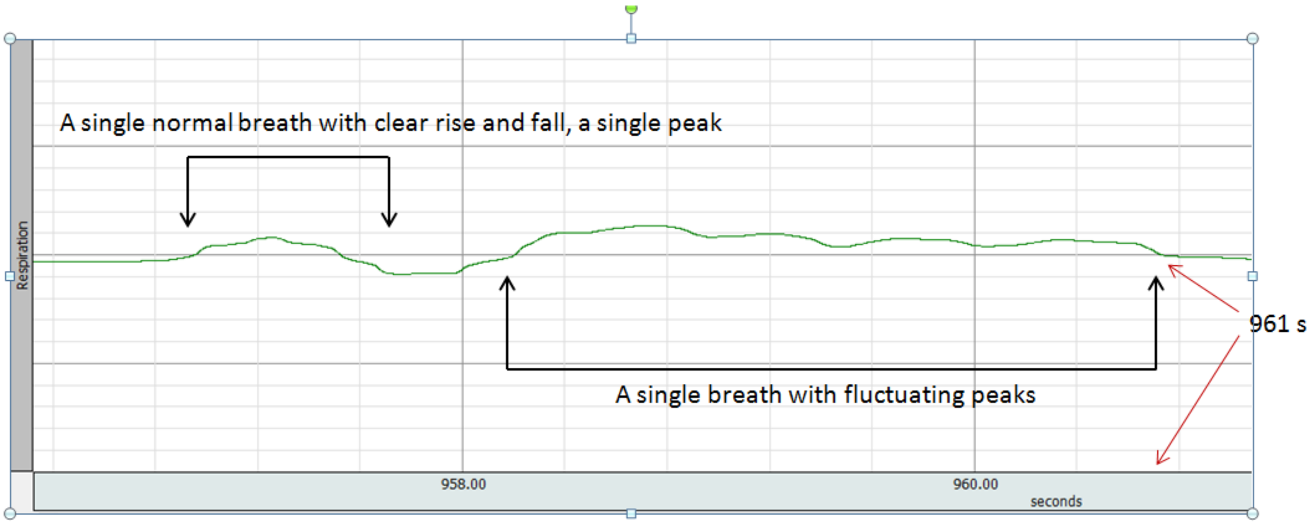

Figure 5: Example of Single Breath with Fluctuating Peaks

The figure above shows how certain distinctions between single breaths were made in this experiment. The first breath has a clear rise and fall, with a singular peak, denoting a single breath. The second breath has an extended plateau, with several similar and fluctuating peaks, but does not clearly fall until much later. The single breath is thus determined by the fall of this plateau, which occurs at 961 seconds in this diagram. This is the method used across all respiratory rate calculations in this experiment.