Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

A

Massey University
Library

The After-Effects of Physical Exertion on Cognitive Performance: Youth Sailors and Logical Reasoning

Gavin R. Dagley

This thesis is presented in partial satisfaction of the requirements for the degree of
Master of Arts at Massey University, Albany
1999

Abstract

Youth sailors completed four logical-reasoning test administrations to explore whether an extended-duration cognitive task might isolate a cognitive performance decrement after long-duration moderate-intensity physical exertion. The cognitive task, the 30-minute Grammatical Transformation Test (GTT), was developed from the three-minute Baddeley Logical Reasoning Test (Baddeley, 1968). Thirty-seven 16 to 20 year-old sailors completed the GTT before and after a day of on-the-water training, and a day of simulated racing ashore. Although a subjective measure indicated the sailors were more physically tired after a day of sailing than after either a day ashore, or prior to either activity, no significant test performance changes occurred due solely to the after-effects of the physical exertion. However, the number of questions correctly answered was higher on the second testing day than on the first, and higher in the afternoon administrations than in the mornings. Further, the number of correctly answered questions increased in each of the consecutive 10-minute test phases in the mornings but decreased in the afternoons. The accuracy of the responses in the first 10-minute phase of each test was lower than in the second and third phases. Also, the accuracy of responses to relatively simpler questions was lower than for complex questions. Finally there were interactions affecting the accuracy of responses between the question complexity and test time, complexity and phase, and complexity, day, time, and phase. The results are discussed in relation to a model of the relationship between fatigue and cognitive performance, and to the cognitively complex sport of sailboat racing.

Acknowledgements

I would like to extend my gratitude to Dr Jennifer Stillman for her support and effort in helping me complete this thesis. I have learned that a supervisor's job is not an easy one. Her persistence and skill is reflected in the consistent improvements that have occurred with each successive draft. I would like to thank Dr Philip Voss, Dr Mark Andersen, and Dr Harriet Speed for their patience with my outbursts of enthusiasm, and for their sound direction. Richard Mason and Harold Bennett of the Royal New Zealand Yacht Squadron were pivotal in allowing this study to occur at all, and I would like to acknowledge their assistance and enthusiasm. I am deeply grateful for the help and friendship of John Weston, youth coach at the Royal New Zealand Yacht Squadron, and previously, the Bucklands Beach Yacht Club. John, as is typical of him, went above and beyond the call of duty to help me through this project. To my family and friends, thank you for your patience and support. Finally, I am deeply indebted to the youth sailors of the Royal New Zealand Yacht Squadron and the Bucklands Beach Yacht Club for their effort in completing a series of unappealing tests, simply because they were asked too.

Contents

Abstract ii

Acknowledgements iii

Contents iv

List of Tables vi

List of Figures vii

1 Background 1
- 1.1 Fatigue 1
- 1.2 Fatigue from Physical Exertion 5
- 1.3 Fatigue from Cognitive Exertion 11
- 1.4 Sleep Loss and Circadian Rhythm Disruption ✓ 12
- 1.5 Points of Difference in Fatigue Research 19

2 Introduction 21
- 2.1 Hypotheses and Theories of Cognitive Performance 21
- 2.2 Long-Duration, Moderate-Intensity Exertion 25
- 2.3 Cognitive Task 26
- 2.4 Factors Affecting Cognitive Performance 31
- 2.5 Study of the After-Effects of Physical Exertion On Cognitive Performance 32

3 Grammatical Transformation Test 36
- 3.1 Development 36
- 3.2 Pilot Study 38

4 Method 41
- 4.1 Participants 41
- 4.2 Materials 41
- 4.3 Procedure 42
- 4.4 Data Analysis 45

5 Results 46
- 5.1 Data Screening 46
- 5.2 Tiredness and Effort Ratings Analysis 49
- 5.3 Total Correct Data Analysis 51
- 5.4 Proportion Correct Data Analysis 53

6 Discussion 63
- 6.1 Overall Performance 63
- 6.2 Simple Versus Complex Question Responses 67
- 6.3 Practice and Time-of-Testing Effects 69
- 6.4 Comparison with Baddeley Test 71

6.5 Model of Fatigue and Cognitive Performance 73
6.6 Implications for Racing Sailors 74
6.7 Problems and Future Directions 75

References **77**

Appendix A **81**

Appendix B **83**

List of Tables

Table 1.1 Points of Difference in Studies of the Effects of Fatigue on Cognitive Performance 4

Table 1.2 Differences in Physical Exertion Types, Durations, and Measures 20

Table 2.1 Cognitive Task Durations in Studies of the Effects of Fatigue on Cognitive Performance 28

Table 4.1 Program for Sailing Days 43

Table 4.2 Program for Lay-Days 44

Table 5.1 Means and Standard Deviations for Physical Tiredness, Mental Tiredness, and Test Effort Subjective Measures on a 7-Point Likert Scale 49

List of Figures

Figure 2.1 Model of relationship between fatigue-inducing factors and cognitive performance. 32

Figure 5.1 Interaction of day and time of test on physical tiredness. 50

Figure 5.2 Interaction of 10-minute phase and time of test on number of questions correctly answered in the GTT. 52

Figure 5.3 Interaction of chronological test day and time of test on number of questions correctly answered in first three minutes of GTT. 54

Figure 5.4 Interaction of chronological test day and time of test on response accuracy proportions in first three minutes of GTT. 57

Figure 5.5 Interaction of question type and time of test on response accuracy proportions in the GTT. 58

Figure 5.6 Interaction of question type and 10-minute test phase on response accuracy proportions in the GTT. 59

Figure 5.7 Interaction of question type and phase on response accuracy proportions in the GTT on the sailing day morning administration. 60

Figure 5.8 Interaction of question type and phase on response accuracy proportions in the GTT on the sailing day afternoon administration. 60

Figure 5.9 Interaction of question type and phase on response accuracy proportions in the GTT on the lay-day morning administration. 61

Figure 5.10 Interaction of question type and phase on response accuracy proportions in the GTT on the lay-day afternoon administration. 61

1. BACKGROUND

When a person declares that he or she feels tired, that person is reporting the subjective experience of "fatigue". Although the experience of tiredness is common enough, definition of fatigue as a psychological, rather than physiological, construct has proved to be problematic. Difficulties have arisen from two primary causes; the large number of variables that appear to influence the experience of fatigue, and inconsistent findings regarding the effects of fatigue on psychological test outcomes. In considering research that has sought to establish a predictive relationship between fatigue and human performance, Holding (1983) commented that there were "very few successful studies" (p155). He quotes studies from as early as 1937 that have shown little or no decrement in operator cognitive and motor performance, even in some quite extreme circumstances.

In attempting to understand the relationship between fatigue and performance, an exploration of the definition of fatigue is essential.

1.1 Fatigue

A dictionary definition of fatigue is "tiredness from physical or mental effort" (Webster's Universal Dictionary and Thesaurus, 1993, p. 206). Holding (1983) extends this already broad definition to include the effects of sleep loss by defining fatigue as "all the consequences resulting from deprivation of rest" (p. 145). The former definition describes fatigue as a psychological construct; a subjective condition resulting from exertion. The latter definition portrays fatigue as "consequences"; that is, all the outcomes from loss of rest, including the subjective perception of tiredness. These outcomes can be distinguished.

Fatigue has been described as having objective, physiological, and subjective components (Bartley, 1976; Holding, 1983). The objective component can be measured as the reduction in work output. The physiological component is represented by changes at tissue level. "Tiredness", however, is a feeling, and represents the experience of fatigue. That is, the subjective component of fatigue is tiredness, depending on whether a particular author included a subjective component in her/his definition of fatigue. Although this three-part construction tends to improve one's understanding of fatigue, there are some difficulties for a psychological researcher in such a division.

To define fatigue in terms of work output would render research into the relationship between fatigue and output circular. Furthermore, work output is not intrapersonal. That is, fatigue is an internal condition that may influence output, whereas work output is external to the person (Bartley, 1976), and may be affected by a variety of factors, including physiological impairment, and psychological and environmental factors.

Although fatigue can be distinguished from its outcomes, it can also be distinguished from arousal. Tomporowski and Ellis (1986) reviewed studies that explored the relationship between fatigue from physical exertion and cognitive performance. They proposed two explanations for the discrepant findings between the studies in the review. One of the explanations concerned the distinction between arousal and fatigue (the other concerned motivational factors). They argued that physical exertion simultaneously arouses the central nervous system and physically fatigues the musculo-skeletal system. This is supported by arousal theories such as the inverted U hypothesis, and the drive-theory hypothesis, which both predict initial improvements in cognitive performance as a result of physical arousal.

Fatigue can also be distinguished from other internal factors that may affect physical and cognitive performance. Wendt and Palmerton (1976) discussed the distinction between subjective fatigue and motivation. They argue that these constructs can be distinguished in terms of attitudinal characteristics and chronobehavioral functioning. That is, motivation can be distinguished by the propensity to expend effort irrespective of available internal resources to support such expenditure. Fatigue, on the other hand, relates to an evaluation of the amount of perceived resource available to meet demands for effort. Wendt and Palmerton (1976) support this argument with evidence of conflicting patterns of motivation and fatigue levels over time. They found that, during the course of a day, fatigue measures followed a regular pattern of highs and lows, while the variation in motivation did not seem to follow any particular diurnal pattern.

Fatigue has also been distinguished from expectancy, sensory habituation and adaptation, boredom, and attitude to risk and effort (Hockey, 1983). Nevertheless, distinguishing fatigue from other psychological constructs has not provided a clear definition of what fatigue *is*. The principle difficulty appears to be in reaching agreement on a sufficiently precise definition for a term that has been used to describe a very broad range of human conditions and performance effects. This has not been resolved in the research.

Although "fatigue" has been demonstrated to result in changes in objective, subjective, and physiological measures, the construct of fatigue can be distinguished from each. From as early as 1921 there have been suggestions to abandon the term fatigue and talk only of the effects of exertion or deprivation of rest (Holding, 1983).

An alternative to the definition of fatigue as a construct, is a classification of fatigue research, with a view to identifying the classes of determinants and effects of fatigue. Bartley (1976) considered that studies "given the fatigue label" (p. 413) could be classified into four groups: those dealing with physiology, those exploring changes in human body functioning (such as changes to the visual and auditory systems), those focusing on measures of physical exertion, and those considering intellectual performance. These groupings are based on the outcomes of fatigue.

It is the last of these outcome groupings that is the focus of the present study; specifically the effects of fatigue on cognitive performance. Studies that fall within this outcome-based grouping can be further sub-divided according to the method used to induce the fatigue. These methods are physical exertion, cognitive exertion, sleep loss, and circadian rhythm disruption. For the purposes of the present study, the term "fatigue" is used to describe the after-effects of exertion (physical or cognitive), sleep loss, or circadian rhythm disruption on an individual. (A subjective evaluation of "tiredness" is also used and will be discussed in more detail later.) The points of difference between individual studies of the effects of fatigue on cognitive performance are described in Table 1.1. The studies themselves are discussed in the following sections.

Table 1.1

Points of Difference in Studies of the Effects of Fatigue on Cognitive Performance

Point of Difference
naturalistic versus laboratory environments
types of fatigue induction: sleep loss circadian rhythm disruption physical exertion cognitive exertion
cognitive performance tasks, constructs and measures: attention / concentration memory learning reaction time arithmetic calculation decision making motor skills flying a plane / driving a truck or train multi-test batteries general intelligence controlled versus automatic tasks
cognitive test methodologies: time of testing in relation to fatigue induction concurrently immediately after induction after a recovery period duration of testing brief (less than 20 minutes) extended (longer than 20 minutes)
individual difference moderators: physical fitness age menstrual cycle
outcomes: performance increases performance decreases performance increases and decreases no change

1.2 Fatigue from Physical Exertion.

The studies that have attempted to establish a predictive relationship between physical exertion and cognitive performance have produced no clear relationship. Davey (1973) described two studies in which participants who engaged in minimal to severe physical exertion, demonstrated significant differences in cognitive performance depending on the level of exertion. The first study utilised three participant groups. The experimental group engaged in two minutes of demanding pedalling on a bicycle ergometer, while the first control group performed a tapping task, and the second control group was rested. The experimental group performed significantly *better* on an immediately subsequent test of auditory vigilance and short-term memory, than either of the two control groups.

Developing this line of research, Davey (1973) conducted a second study that varied the amount of exertion for each of five groups. The four participants in each group pedalled on the ergometers for either 15 or 30 seconds, or 2, 5, or 10 minutes. Davey (1973) found that the moderate exercise groups (30 seconds and two minutes) showed a significant performance improvement, while the extreme exertion group showed a significant performance decrement. He concluded that there appeared to be "an 'inverted U' relationship between physical exertion and subsequent mental performance" (p. 598) and that this relationship may be due to the effects of exertion on physiological arousal.

Two years previously, Hammerton (1971, cited by Holding, 1983) had conducted a series of three studies exploring the effects of demanding exercise on cognitive performance. Participants were required to complete a strenuous stepping task (a version of the Harvard step task, one foot up and down each second) for 200 seconds, and, as a measure of cognitive performance, complete a grammatical transformation task. The various studies sought to measure performance on the cognitive task using a range of presentation times and methods. No performance differences were found between the "fatigued" group and the rested control group when the cognitive test was completed immediately after the physical exertion (or under any other administration method).

Although the comparability of the Davies (1973) and the Hammerton (1971) studies could be argued, similar, apparently conflicting, results appear to characterise much of the research on the cognitive effects of fatigue from physical exertion. This was apparently the view of Tomporowski and Ellis (1986) when they undertook their review of the area.

The purpose of their review was to consider whether different physiological statuses, caused by physical exertion, led to different effects on cognitive functioning. After discussing the existing theories in the area, Tomporowski and Ellis (1986) proceeded to describe each of the 26 reviewed studies in terms of methodological and outcome factors, and to classify the studies in terms of the type of physical exertion.

The physical exertion classification was based on the duration, intensity, and aerobic or anaerobic nature of the exercise. This resulted in four categories; (i) very brief, high-intensity anaerobic, (ii) short-duration, high-intensity anaerobic, (iii) short-duration moderate-intensity, aerobic, and (iv) long-duration aerobic exercise. Each study within a category was then described in terms of: the measure of preliminary fitness of the participants, the nature of the physical exertion and the cognitive task used to measure performance, the timing of the cognitive task administration in relation to the physical exertion, and the results. This process highlighted the wide range of methods used to induce exertion. The very brief anaerobic tasks used various means of achieving the physical resistance that resulted in the need for exertion, primarily through the use of weights or springs. The longer duration tasks tended to use some form of sustained leg exercise. This amounted to some kind of walking, running, cycling or climbing.

More importantly, a wide range of cognitive measures was used. These tasks included figure perception, memory tasks, complex reasoning, signal detection, sentence verification, arithmetic tasks, and perceptual organisation. These tasks were assessed in terms of both reaction time and accuracy. Overall, only two studies obtained results indicating general impairment on the cognitive task as a result of physical exertion, while two further studies showed impairment of low fitness participants only. Eleven of the studies showed a facilitation of cognitive performance resulting from the exercise, while results from another two studies indicated an initial facilitation of cognitive performance, but with a performance decrement after 10 minutes or more physical exertion.

Tomporowski and Ellis (1986) concluded that no general statement could be made about the effects of exercise on cognition. There did, however, appear to be some patterns based on exercise duration and intensity. Firstly, in the case of very brief duration exercise, moderate levels of exertion appeared to facilitate cognitive performance, while very low and very high levels tended to have no effect. This was supported by the studies that, with only one exception, indicated facilitation of cognitive performance. Secondly, short-duration aerobic exertion tended to impair or have no effect on cognitive performance, and fitness may have interacted with exertion to produce some variation in the effect. Thirdly, long-duration exertion generally facilitated cognitive performance. Finally, short duration high-intensity exertion tended to provide inconclusive results with respect to cognitive performance.

Tomporowski and Ellis (1986) offered two possible explanations for these results. Firstly they differentiated between the constructs of arousal and fatigue arguing that these operate independently and, at times, inversely on cognitive performance. They believed that the cognitive performance differences related to fitness might have been due to exertion affecting fit and less fit participants differently. Secondly they argued that motivation had had a substantial influence on test performance. Although the authors were referring to the differences in motivation between exercisers and non-exercisers on the physical exertion task, this motivational difference is likely to apply equally to performance in the cognitive tasks.

Tomporowski and Ellis (1986) drew some conclusions from their review. Firstly, they concluded that the lack of a theoretical base has resulted in a piecemeal approach to the research. Secondly, much of the difficulty in gaining an understanding of the effects of physical exertion on cognition was due to a lack of consistency in the methodologies employed. Tomporowski and Ellis (1986) also made recommendations regarding future research. They considered that such research should be theory-based, and should control for the effects of physical fitness and motivation. Likewise, future studies must consider the nature, timing, duration, and intensity of the exertion, and attempt to measure the physiological effects thereof.

The Tomporowski and Ellis (1986) review highlights some interesting points. Firstly, the review highlights the age of some of the research in this area. Though the article was published in 1986, only two of the reviewed studies were dated 1980 or later.

More than half of the 26 articles were dated prior to 1970. Secondly, the study did not consider other factors that may have affected the performance outcomes. As previously discussed, Tomporowski and Ellis (1986) described the studies in their review in terms of five factors; participant fitness, exertion intensity and duration, the nature of the psychological task, the point of cognitive measurement, and the performance outcome. Factors that could equally have affected results include the duration of the psychological testing, and the "controlled" or "automatic" nature of the cognitive tasks. Such factors have been considered in research on the cognitive effects of sleep loss and circadian rhythm disruption, and are discussed later.

In summary, Tomporowski and Ellis (1986) found that there was considerable inconsistency in the results of studies exploring the effects of physical exertion on cognition. Tomporowski and Ellis (1986) stated "we believe that the data obtained from the studies we reviewed fail to provide clear support for the notion that exercise influences cognition" (p. 345). They considered that confounding motivational variables, and the interaction of the separate constructs of arousal and fatigue, were important factors in generating the inconsistent results. Factors that were not considered by Tomporowski and Ellis (1986) were the nature ("controlled" or "automatic") and duration of the cognitive tests employed.

Research published subsequent to the Tomporowski and Ellis' (1986) review seems only to add to the equivocal nature of the results in the area. These later studies can be broadly classified in terms of the intensity of the exertion task - moderate or severe.

McNaughten and Gabbard (1993) completed a study of moderate intensity exercise that indicated a cognitive performance improvement. This study used walking as the exertion task, and measured the effects of 20, 30, and 40 minutes walking at moderate intensity on the cognitive performance of 120 sixth-grade boys and girls. The cognitive test consisted of 90 seconds of speeded arithmetic computation. The investigation used a Solomon Four-group design. From a comparison of the results of each group, the researchers found that there was a significant improvement in test scores resulting from an interaction between the level of exertion and the time of day the tests were administered. The 30- and 40-minute walks later in the day produced an improved test performance. There were no significant main effects of level of exertion, duration of exertion, or gender. The researchers commented that, while the study did indicate a plateau in performance, there was no indication of the decline required to support the inverted U theory of arousal and performance. This may not have been entirely unexpected given that the exertion was limited to "moderate intensity" (p. 1156).

In a study of more generalised physical exertion, Raviv and Low (1990) compared groups of junior high school children to test teachers' contentions that physical education classes negatively effected children's concentration. Using a brief, but intense test of concentration (estimated duration approximately 5 minutes), the children were tested at the beginning and the end of physical education and science classes. The test was completed under time pressure, and required the students to locate all the occurrences of a complex target character from within lines of similar characters. Performance was assessed on the basis of number of characters correctly located, less the number incorrectly marked, and on the basis of accuracy. Although the study did indicate performance differences between the beginning and the end of class, and differences related to time of day, no differences were found between the physical education and science class groups.

Using a between-subjects design, Sparrow and Wright (1993) tested the effects of moderate physical exercise on separate groups of 10 physically active men of equivalent fitness. Two further groups acted as control, and either played Bingo or had no activity. Using three levels of a six-minute bench stepping exercise as the exertion task, the groups were tested prior to, and subsequent to the exertion or control conditions. The researchers used an eight-minute session of Raven's Progressive Matrices and a variation of the WAIS Arithmetic test as the cognitive measures, and the results were compared across the groups. No cognitive performance differences were found between groups. However, the low power afforded to a between-subjects design with 10 participants in each group may have influenced these results. The researchers also found no interaction between arousal and fatigue effects on cognitive performance, as suggested by Tomporowski and Ellis (1986). The researchers argued that the homogeneous level of the participants' fitness, and the variation in exercise levels should have allowed such an interaction to occur. Such an assertion assumes that effects of arousal and fatigue will not cancel each other out at each of the quite moderate exertion levels of the study.

There have also been a number of studies of the effects of more extreme physical exertion. Hancock and McNaughton (1986) studied a group of six experienced orienteers who were assessed on a number of cognitive functions while engaged in running on a treadmill. After 15 minutes of warm-up, the runners completed a three-minute burst that took them up to their anaerobic threshold. During this three-minute burst, they verbally responded to a range of questions regarding projected slides of orienteering checkpoints. Given the low number of participants and trials (six orienteers participated, each completing the 100 question test twice), it was surprising that the researchers were able to find a significant decrement in performance during the exertion, when compared to the same test performed at rest. (The order of testing conditions was counterbalanced to control for practice effects).

Soetens, Hueting, and Wauters (1992) used a visual perception task to test the cognitive effects on 20 participants of an exhausting session on a bicycle ergometer. Thirty minutes after completing the 30-minute cycle at progressively increasing loads, the participants were required to complete a 15 minute trial in which they reported the number of dots (between 3 and 12) on a series of cards presented tachistoscopically. Both accuracy and reaction time were recorded. A significant decrement in accuracy was found compared to the same participants performances in a rested condition (order counterbalanced), particularly for cards which had a large number of dots. No differences were found in reaction times.

Zervas (1990), in a between-subjects study of 374 men and women, found no effect from 40 minutes of aerobic exercise on a 20-minute speeded version of the Eysenck test of intelligence. A further study by Zervas, Danis, and Klissouras (1991) found no difference on a test of cognitive performance between monozygotic twin boys (aged 11 to 14) after one of the pair completed a physical training program over a period of 25 weeks. However, both siblings (fit and unfit) displayed an improvement in reaction time and in accuracy on a design-matching test, after completing 20 minutes on a running treadmill at an intensity above their respective anaerobic thresholds. The participants were tested for a period of 15 minutes (including two 2.5-minute breaks) both prior to, and 15 minutes after, completion of the treadmill task.

Féry, Ferry, Vom Hofe and Rieu (1997) sought to clarify the effect on cognitive performance of maximal exertion until exhaustion. Using a within-subjects design, the researchers required 13 fit men to pedal on a cycle ergometer in two counterbalanced conditions. The first condition was a three-stage progressive load session with the third stage requiring participants to pedal until exhaustion at 90% of maximal aerobic capacity. The second condition used a constant load that required exertion at the lowest level of the progressive condition and provided the control data. Each trial in the cognitive test required participants to determine whether a visually presented consonant was present in a previously displayed group of 4 or 7 consonants. Response time was measured. Testing was conducted at the lowest and highest exertion levels, and after exertion. Each test lasted 14 seconds with reaction time as the dependent variable. The only significant result was an interaction between extreme exertion levels and the more complex elements of the cognitive task. That is, cognitive performance decremented under extreme exertion combined with the more complex seven-consonant trials.

It is possible that the different results obtained by Féry et al. (1997) and Zervas et al. (1991) are attributable, among other factors, to the 15 minutes of rest allowed in the Zervas et al. study, prior to the post-exertion test. Using Tomporowski and Ellis' (1986) two-factor explanation, some recovery from the fatigue component may have occurred without a similar reduction in the arousal component.

In summary, research since Tomporowski and Ellis' (1986) has complicated rather than clarified that review's findings. That is, the effects moderate duration, high-intensity exertion have tended to show decrements in cognitive performance (with the exception of Zervas 1991), while more moderate intensity exertion seems to produce equivocal results; sometimes an improvement and sometimes no effect.

1.3 Fatigue from Cognitive Exertion.

Only a limited amount of research has been carried out on the effects of cognitive fatigue on a test of cognitive performance. A general discussion of fatigue by Holding (1983) did not specifically address cognitive fatigue per se, but considered the effects of perceptual fatigue, particularly as it relates to vigilance. He considered that one of the factors affecting vigilance was the general level of arousal. By way of example he considered that sleep loss would reduce arousal and signal density would increase arousal.

Rust and Barnard (1979) administered the Performance and Verbal scales of the Weschsler Adult Intelligence Scale to a group of 44 elderly men and women in counter-balanced order. They sought to determine whether decreases in the Performance Scale scores for older individuals were a result of the fatiguing nature of the task itself. They found no significant differences in performance as a result of the order the tests were given.

Sothmann, Hart and Horn (1992) sought to determine whether a program of exercise which improved aerobic fitness, would effect the sympathetic nervous system response to stress. A group of 24 previously sedentary men undertook an aerobic exercise program, and completed a cognitive test battery in both the pre- and post-exercise conditions. The stress variable was the administration of 18 minutes of a modified Stroop test (naming the colour of a word where the word is the name of a colour other than that colour in which the word is displayed). Both before and after completion of the Stroop test, participants were required to complete a series of anagrams and the time taken to complete the task was recorded. The researchers found that, although the exercise program had no effect on anagram completion times, the participants took significantly longer to complete the anagram task after the Stroop test than before that task. There is therefore an indication that, in some circumstances, cognitive performance may be affected by completing an immediately preceding cognitive task.

1.4 Sleep Loss and Circadian Rhythm Disruption.

It seems clear that factors in addition to those identified by Tomporowski and Ellis (1986) are at work in the relationship between physical exertion and cognitive performance. A related area of fatigue research, sleep loss and circadian rhythm effects, may shed some light on these other factors.

The area of sleep loss and circadian rhythm disruption has held particular interest for shift work researchers. In 1995 the United States based National Transport

Safety Board (NTSB) and the NASA Ames Research Center jointly hosted a symposium on fatigue. Chief of the Operation Factors Division of the NTSB, Danaher (1995), remarked on the detrimental implications of irregular work and rest cycles and cited a series of examples of transport accidents in which fatigue was implicated. Examples were provided from rail, sea, and air transport as testimony to the serious consequences of fatigue occurring in critical operations such as vehicle control. Interestingly, Danaher pointed out that it was not until 1993 that the NTSB first cited fatigue as a causal factor in an air carrier accident.

Rosekind et al. (1992) provided an overview of the NASA Ames Fatigue Countermeasures Program. The objectives of the program were to determine the extent of fatigue in its various forms on flight operations, to determine its effects on flight crew performance, and to develop a countermeasures program to minimise the adverse effects on performance. In the course of the research, the program has used subjective, physiological, behavioural, and cognitive performance measures to identify the presence of fatigue, and to measure its effects on performance. These measures have also been used to assess the effectiveness of countermeasure strategies. One of the primary conclusions drawn from this extensive program of study is that sleep loss and circadian rhythm disruption from long haul operations can result in a measurable increase in fatigue and sleepiness as indicated by physiological and performance measures.

NASA and Air New Zealand have formed a Flightcrew Fatigue Study Group that has worked collaboratively over a number of years. One outcome of this effort is the study by Dawson and Petrie (1996) on flightcrew fatigue on a long haul tour of duty from Auckland to Frankfurt and return. The study measured physiological, subjective, and performance variables, to ascertain the pattern of fatigue and sleep loss during the tour. The study reported two interesting findings. Firstly, in confirmation of previous research, the writers found that measuring the pattern of fatigue and sleep loss was possible without the presence of researchers accompanying the pilots. Secondly, and more controversially, the writers found a strong association between the subjective and performance measures of fatigue. This finding was contrary to those of a number of writers (Blagrove, Alexander & Horne, 1995; Dinges, 1990; Holding, 1983; and Krueger, 1989). In explaining the inconsistent relationships they found between subjective fatigue and performance, these writers suggest that subjective fatigue is influenced by factors other than purely internal physiological fatigue states. Examples of such other factors are workload and external stimulation. These are factors that would have been expected to influence the Dawson and Petrie (1996) study. Dawson and Petrie make no comment as to the reason for this unusual finding. Nor do they comment on the possibility of distortions in the self-report data due either to participant knowledge of their results on the performance measure, or to an expectation on the part of the participant as to what the subjective rating should be at any given time.

A large number of studies outside the aviation industry have investigated the effects of sleep loss. These studies have distinguished between the effects of sleep deprivation (the preclusion of sleep for an extended period) and sleep reduction (the reduction of diurnal sleeping hours over a period of days or weeks). Lingenfelser et al. (1994), using a within subjects experimental design, evaluated the effects of a night on call on 40 resident physicians. This amounted to partial sleep deprivation. Results from a battery of six tests of performance on routine medical tasks indicated a significant deterioration on all six tests after a night on call. The researchers commented that some of the negative effects of such deprivation may require prolonged testing to isolate, beyond the 25 minutes required to complete the battery of tests they had used. The length of the test battery in their study was limited by practical considerations.

In a correlational study of medical residents, Jacques, Lynch and Samkoff (1990) also found a relationship between the amount of lost hours of sleep in the night prior to testing and performance in a four hour examination. That is, the less sleep the participant had in the night before the examination the worse the performance. Jacques et al. (1990) investigated the effect by comparing the results of 353 Pennsylvania practice residents in their in-training examination. The relationship between hours awake the night before the exam and performance in the exam might also have been a function of how well prepared the participant was. So although the relationship between hours of sleep and performance might be strong, the poor showing of those with less sleep might be due to causes other than sleep loss. Nevertheless, Jacques et al. (1990) characterised the performance difference between a participant with a normal night's sleep and one without sleep the night before the examination as being the same as the performance difference between a first-year and a third-year trainee. Jacques et al. (1990) reported that the average working week for residents was 74.2 hours, with some specialities working, on average, more than 100 hours per week. Thus fatigue may be a significant factor in job performance.

Bohnen and Gaillard (1994) studied the effect of one night of sleep deprivation on 15 male university students. Using a within subjects experimental design, the participants completed two cognitive tasks concurrently for a period of 30 minutes. The order of the control (no sleep loss) and experimental (one night without sleep) conditions was counterbalanced. Using computer terminals, participants were required to maintain a dot within a prescribed area on the screen, while concurrently estimating the passage of time in consecutive 2- and 3-minute intervals for the 30-minute duration of the task. The researchers found that the tracking task was effected by sleep deprivation, particularly towards the end of the 30-minute testing period, while the time estimation task remained unaffected. The researchers discussed the motivational differences between the tasks and concluded that the time estimation task was both less monotonous and more cognitively demanding than the tracking task. They further concluded that, consistent with previous research, the difference in performance between the two tasks after sleep loss was due to the level of inherent motivation provided by each task, the more demanding task providing the greater motivation. The researchers further commented that the "relatively short period of 30 minutes" (p. 1029) may not be long enough to detect changes in performance as the motivational effect of the task starts to wane.

May and Kline (1987) undertook a study of sleep deprivation on 135 non-commissioned officers in a sustained operations field environment. The performance of the soldiers was measured on a battery of 15 three- to seven-minute duration cognitive tests in three conditions: a baseline condition, a physical exertion condition, and a sleep deprivation condition. Five of the available 15 tests were administered in any single test session, with a total testing time of less than 30 minutes per session. After controlling for practice effects and the after-effects of physical exertion, the researchers found a pattern to the decrements in the cognitive test performances when the participants were sleep deprived. They found that the tasks that required a greater cognitive demand showed no decrement, while the less demanding tasks showed significant performance degradation. This was in spite of problems with the administration of the tests and questions concerning the tests' psychometric properties, that would have tended to reduce the probability of finding any decrements in performance.

1.4.1 Use of the Baddeley Logical Reasoning Test.

Linde and Bergstrom (1992) undertook a study of 16 undergraduates deprived of sleep for one night. The researchers used a between-subjects experimental design to compare participants' performance on a battery of four cognitive tests: the three-minute Baddeley Logical Reasoning Test (Baddeley, 1968), the Raven's Progressive Matrices (40 minutes in duration), the Digit Span test (approximately three minutes), and a test of attentional power (of approximately 20 minutes duration). The only test to demonstrate an effect was the Raven's Progressive Matrices, which showed a reduction in the number of correctly answered items after sleep loss. In discussing the "no effect" outcome of the Baddeley test, the researchers commented that they may have been unable to detect a change in performance because the test was too short, and there was a 50% chance of achieving a correct answer simply by guessing ("true" or "false").

The Baddeley Logical Reasoning Test (Baddeley, 1968) is a three-minute speeded test in which respondents determine the accuracy of up to 64 reasoned statements. A sample statement is "A follows B – AB". The correct answer is for the respondent to check "false". The test was developed as a short test of reasoning, and has demonstrated a correlation (+.59) to the British Army verbal intelligence test (Baddeley, 1968).

A series of seven studies by Ryman, Naitoh, and Englund (1985) on the effect of sleep loss and physical fatigue in a sustained operations environment also used the Baddeley Logical Reasoning Test. The researchers administered a computerised version of the Baddeley over a period of five days for each of the seven studies. The studies employed 50 pairs of Marine enlisted personnel, half of whom were put through a series of demanding continuous workdays (walking on a treadmill for up to 17 hours per day), and half of whom rested. Factors that were varied for all participants were the rest type (rest, nap or sleep), length of the rest period (3,4 or 8 hours), and the start time of the various workdays. The researchers were particularly interested in performance differences related to sentence complexity that resulted from either sleep loss or physical exertion.

Ryman et al. (1985) argued that accuracy, rather than the total number correct, was the more important component in what was intended to be a test of "higher mental processes" (p. 1182). The researchers were particularly interested in comparing the response accuracy for the more complex, passively worded statements with that of the less complex, actively worded statements. Although this focus is consistent with research and theory concerning fatigue and task complexity (Shiffren & Schneider, 1977; Tomporowski & Ellis, 1986), the researchers do not clarify how they controlled for the motivation to complete as many questions as possible in a speeded test. Such motivation is likely to affect participant accuracy, and this may be particularly important to the results given the between-subjects design.

Ryman et al. (1985) found that there were significant differences in performance (as measured by accuracy) on the eight different sentence types when comparing the baseline measures to the continuous workday measures. In particular, the more complex passively worded sentences (which had the poorest comprehension rates in the baseline condition) were unaffected by fatigue, while participants demonstrated a significant performance decrease with the less complex actively worded sentences. Further, the study showed that these effects were consistent across the control and exercise groups (who experienced the same variations in rest but not in exercise), and therefore the performance changes were a result of sleep loss and not fatigue from physical exertion. The researchers explained these findings as resulting from the increased attention demanded by the more complex statements.

Blagrove et al. (1995) conducted a series of studies that compared the effects of sleep reduction with those of sleep deprivation. Using a group of three cognitive tests, including an extended version of the Baddeley, the researchers found that one night of sleep deprivation affected performance on all the cognitive measures. In contrast, those participants who reduced their hours of sleep to 5.3 hours of sleep per night or less, showed a consistent decrement only in an embedded figures task. One group also showed reduced performance in an auditory vigilance task, despite subjective reports of severe sleepiness and reduced concentration.

1.4.2 "Sustained Operations" Review.

Krueger (1989) conducted a review of the issues relating to sustained operations with particular emphasis on how factors such as sleep loss and fatigue affect worker stress and performance. Sustained operations consist of long continuous work periods without rest. Krueger provided examples demonstrating that long, uninterrupted periods of work result in reduced performance in a number of cognitive domains. Further, reduced or fragmented sleep over a longer period also reduced performance in these areas. That is, a number of the studies reviewed by Krueger demonstrated cognitive performance decrement as a result of sleep reduction, although such studies tended to produce more conflicting results than did sleep deprivation studies. The decrements were attributed to fatigue resulting from sleep loss and workload, and to the effects of circadian rhythms. A number of factors were shown to reduce the effect of both sleep loss fatigue and circadian rhythm effects. These included task factors such as novelty, duration, difficulty and complexity; and factors that effected participant effort such as knowledge of results, anticipation of a nap, exhortation to perform, and other motivational factors.

A study by Horne and Pettit (1985) also demonstrated the power of motivation to overcome the effects of sleep deprivation. Participants were able to suppress the effects of sleep deprivation on performance of an auditory vigilance task for up to 36 hours when a financial incentive to maintain performance was offered. Although the study sample was small (15 participants divided into three groups), the study also demonstrated some of the effects of circadian rhythms on performance.

In summary, sleep loss and circadian rhythm disruption seem to produce consistent effects on both fatigue and performance. However, sleep deprivation seems to produce more consistent performance effects than accumulated sleep loss. The weight of evidence indicates that self-report data are not entirely reliable predictors of the effects of sleep loss on performance. Finally, the length and nature of the performance measure can effect the outcome, as can subjective motivation and resulting effort.

1.5 Points of difference in fatigue research

Individual studies within fatigue research can be distinguished by a series of points of difference. Table 1.1 provided an extensive, but not definitive, list of these factors and the options that have been employed by the studies reviewed thus far. Each class of factor presents options for every study. Therefore, given the number of options under each class of factor, the number of possible permutations is enormous. Equally the existence of this large range of possible combinations may go some way towards explaining the apparent inconsistencies in the published results.

The present study is concerned with the effects of physical exertion on cognitive performance. As noted by Tomporowski and Ellis (1986), studies to date have demonstrated large variances in the methods and operational definitions employed, and in the outcomes achieved. In addition to the factors listed in Table 1.1, variations in the way in which fatigue is induced through physical exertion may also have an impact on the outcome of the study. A list of such factors encountered in the literature reviewed to this point is included in Table 1.2.

Table 1.2

Differences in Physical Exertion Types, Durations, and Measures

Point of Difference
types of physical exertion:
aerobic
anaerobic
fixed work volume versus fixed work to individual capacity ratio
progressive load increase
duration of exertion:
very brief (less than 30 seconds)
brief (30 seconds to 5 minutes)
moderate (5 to 20 minutes)
long (greater than 20 minutes)
physical exertion measures:
heart rate
VO_2 max
task intensity and duration

The purpose of this study is to investigate whether there is a causal relationship between the after-effects of physical exertion and cognitive performance. Before the relationship between these constructs can be reduced to a predictive model, a review of the current theories that attempt to explain performance differences, is appropriate.

2. INTRODUCTION

2.1 Hypotheses and Theories of Cognitive Performance

Studies of the effects of physical exertion on cognitive performance have produced equivocal results. As previously discussed, these results could be affected by a range of methodological factors. The factors listed to this point are: the type and duration of the physical exertion, the nature, timing, and duration of the cognitive performance tasks, and non-random individual differences in the participants (such as physical fitness). In addition there is a range of theories and hypotheses that attempt to explain and predict cognitive performance, and it is these explanations that are now discussed.

In addition to the effects of "fatigue", explanations of cognitive performance can be grouped into three broad categories: arousal theories, cognitive capacity and effort theories, and hypotheses relating to controlled and automatic processes.

2.1.1 Arousal Theories

The relationship between arousal and performance has been extensively investigated with regard to motor and athletic performance. Two early models of the relationship between physiological arousal and motor task performance have been widely discussed. They are "drive theory", and the "inverted-U" hypothesis (Anshel, 1997). Both models predict an initial improvement in performance as physiological arousal increases. However, whereas drive theory describes a linear relationship between arousal and performance, the inverted-U hypothesis predicts that the relationship will be curvilinear, and performance will peak at some level of arousal beyond which further arousal increases will result in decreased performance. There is a further important difference between the two models. Whereas the inverted-U hypothesis predicts a relationship between the level of physiological arousal and performance, drive theory predicts a relationship between general activation from behaviour and the probability of the occurrence of the dominant or habitual response.

Drive theory defines "performance" as the likelihood of the dominant or habitual response occurring. That is, drive theory predicts that, as arousal increases, the best-learned or habitual response is more likely to occur, irrespective of whether or not the response is correct.

Evidence providing some support for both models has been found in research into the relationship between the effects of physical exertion and cognitive performance. The previously referred to study by McNaughten and Gabbard (1993) uncovered an interaction between time of day and exertion on cognitive test performance. This interaction resulted in improved mathematical performance in school children after 30 or 40 minutes walking, as compared to 20 minutes of walking, during two of the three daily walking times. The researchers commented that there was evidence of a performance improvement and a plateau, but no evidence of the decrement in performance required to support the inverted-U hypothesis. In contrast, the previously discussed results of Davey's (1973) research, which employed a range of cycling tasks and a test of memory, provides support for the curvi-linear relationship between arousal and cognitive performance described by the inverted-U hypothesis.

The obvious difficulty in interpreting this research as it relates to arousal is that the studies do not distinguish between arousal and fatigue effects. May and Kline (1987) comment that this is also an issue with research into the sleep loss. The concept of a unidimensional view of arousal has also been questioned. Linde and Bergstrom (1992) discuss distinctions between arousal, activation, and effort, including the view that arousal is perceptual, activation controls motoric responses, and effort has a superordinate function in relation to cognitive performance.

Sport psychologists have had a particular interest in investigating the effects of arousal and anxiety on athletic performance. The distinction between these two constructs is not entirely clear. In contrast to the distinction drawn by Linde and Bergstrom (1992), Kremer and Scully (1994) describe arousal as a term most commonly used to refer to a physiological state. Anxiety is described as more commonly referring to a stress response. Suffice to say that there is considerable overlap between the constructs (Kremer & Scully, 1994).

In investigating the relationship between arousal/anxiety and performance, additional explanations have been proposed. These explanations include catastrophe theory (Kremer & Scully, 1994), and a refinement of the inverted-U hypothesis, which proposes that different motor tasks have different levels of optimal arousal and therefore the inverted-U curve might be task specific (Anshel, 1997). These models also predict an initial improvement in performance, however their predictions diverge when considering the effects of further increasing arousal beyond that which provides this initial performance improvement.

Authors of additional models of the relationship between arousal/anxiety and athletic performance have questioned the unidimensional approach to describing anxiety and the effect of interpretation of anxiety. Kremer and Scully (1994) discuss multidimensional anxiety theory and, in describing the relationship between anxiety and performance, they distinguish somatic responses to anxiety from cognitive responses. The model predicts that somatic responses will have an initially beneficial effect on performance, and that further increases in anxiety will result in a curvi-linear relationship with performance, not unlike that of the inverted-U hypothesis. The model predicts that cognitive responses to anxiety will have a negative linear relationship with cognitive performance.

Reversal theory (Kerr, 1985, cited by Anshel, 1997) and research by Hanton and Jones (1999) consider both the interpretation and the magnitude of anxiety in predicting the effect on performance. Reversal theory is based on the premise that a person can interpret anxiety as either pleasant (e.g., excitement) or unpleasant and that such interpretations affect the emotional state of the person and therefore performance. In addition, these interpretations can alter rapidly. For example, a person who completes a difficult and stressful task will reverse a negative interpretation of anxiety into a positive interpretation, while the magnitude of the anxiety need not change. Hanton and Jones (1999) applied sport psychological interventions to help a group of elite swimmers to re-interpret pre-competition anxiety symptoms as positive. Their intervention resulted in improved swimming performance and subjective interpretations of anxiety, but with no change in the intensity of the anxiety levels.

Tomporowski and Ellis (1986) summarise the theoretical positions concerning arousal as it effects cognitive performance. Firstly, they have limited their consideration of arousal theories to drive theory, and the inverted-U hypothesis. This appears to be consistent with the lack of research that considers alternative explanations. Secondly, they point out the complex nature of the arousal state, and the difficulty in adequately defining or operationalising the condition. Thirdly, they note the lack of distinction between fatigue and arousal effects in cognitive performance research, and that these two conditions may exert opposite influences on cognitive performance.

2.1.2 Cognitive Capacity and Effort

As part of an extensive general theory of the allocation of cognitive capacity according to effort, Kahneman (1973) proposes that individuals have a limited attentional capacity available for information processing, and that, within this fixed limit, physical arousal levels affect available capacity. More difficult and novel tasks result in greater arousal, and, therefore, greater cognitive capacity. Where there are excess demands on cognitive resources, then capacity is allocated according allocation policies that, in turn, are affected by enduring dispositions, momentary intentions, and evaluation of demands on capacity.

However, Kahneman (1973) uses the terms "momentary capacity, attention, or effort" (p.13) as interchangeable terms in the discussion of attentional capacity. This suggests that either "capacity" involves conscious choice, or "effort" does not. Holding (1983) suggests that one of the results of fatigue may be the resulting choice of the individual to make less effort on cognitive tasks. Whether or not effort and capacity are interchangeable terms in this context, the implications for the present study are clear. Effort can be expected to have an impact on performance in cognitive tasks, and relatively more demanding tasks may result in greater effort on the part of the participant.

2.1.3 Controlled and Automatic Processing

Shiffren and Schneider (1977), in their extensive discussion of human information processing, distinguish between controlled and automatic processing. They describe controlled processes as tasks that are demanding in terms of short-term memory and attention, and that can be adopted and modified easily. Automatic processes are tasks that are not demanding in terms of short-term memory or attention, that require considerable training to develop, and are difficult to modify. Controlled tasks, with practice, may become automatic over time. Those tasks that remain primarily controlled in nature will very quickly demonstrate a plateau in performance level.

Shiffren and Schneider's (1977) theory can be distinguished from Kahneman's (1973) work by their exclusion of the concept of effort from their discussion. That is, Shiffren and Schneider (1977), in developing their theory, have assumed that participant effort has been consistently maintained at the highest level. Kahneman (1973), on the other hand, attempts to explain the allocation of cognitive resources on the *basis* of effort.

Soetens, Hueting, and Wauters (1992) argued that the effect of fatigue is a shift from controlled to automatic processes, with the result that the more demanding controlled processes would be disturbed. However, this argument is at odds with the outcomes of studies such as May and Kline (1987), in which the more complex tasks seemed least affected by fatigue. Bohnen and Gaillard (1994) also found a performance decrement in the less demanding of the cognitive tasks (visual tracking) that were completed by their sleep derived subjects. A decrement was not found in the participants' performance on the more demanding time estimation task.

Although the distinction between tasks that require controlled and automatic processing appears to be reasonably clear, the affect of fatigue on tasks with differing processing requirements is unclear and is discussed in more detail below.

2.2 Long-Duration, Moderate-Intensity Exertion

The present study is concerned with the effects of long-duration, moderate-intensity physical exertion on cognitive performance. In the present study, "long-duration" is defined as in excess of 30 minutes. "Moderate-intensity exertion" is defined as physical exertion that, while demanding, is less than the extreme and exhausting demand of maximal effort. These definitions have been constructed from the categories defined by Tomporowski and Ellis (1986).

The three studies that met the "long-duration" criterion in Tomporowski and Ellis' (1986) review ranged in exertion duration from approximately 28 minutes to 4 hours. The exertion tasks were comprised of a marathon race, a five-mile march with a 40-pound pack, and a treadmill run until exhaustion. Unlike the other categories in their study, Tomporowski and Ellis (1986) did not ascribe an intensity indicator to the long-duration group, and the tasks were simply described as "aerobic exercise" (p.341). Two of the studies found facilitative effects from the exertion on the respective cognitive tasks, while the third obtained no effect.

Five further studies were described as "short-duration, moderate-intensity aerobic exercise" (Tomporowski & Ellis, 1986, p. 340) and ranged in duration from 10 to 20 minutes. The exertion tasks consisted of callisthenics, running and walking, and bicycle pedalling. Results from two of the studies indicated that there were performance decrements on cognitive tasks after the physical exertion. Two further studies indicated physical exertion had no effect on cognitive task performance. The results of the fifth study indicated a curvilinear relationship existed between reaction time and duration of exertion.

Only one study has been identified which could be described as having long-duration, moderate-intensity exertion as the independent variable in a study of cognitive performance. This is the previously referred to McNaughten and Gabbard (1993) study of school children where the exertion consisted of 20 to 40 minutes of walking. "Moderate intensity" (p. 1157) exertion was defined as walking while maintaining a heart rate of between 120 and 145 beats per minute. Among the results was an interaction between time-of-day and the length of time spent walking, with longer walking times, later in the day, resulting in improved performance on an arithmetic test.

The effects of physical exertion on cognitive performance in the studies that have used long-duration exertion, and/or moderate-intensity exertion have variously indicated performance improvements, performance decrements, or no effect.

2.3 Cognitive Task

A common feature of the research into the effects of physical exertion on cognitive performance has been the relatively short duration of the cognitive tasks employed. Table 2.1 lists the studies discussed to this point and the durations of the respective cognitive tasks. It could be expected that the effects of physical exertion or sleep loss would be more likely to be evident in longer, rather than shorter tests of cognitive performance. Lingenfelser et al. (1994) commented that detection of the negative effects of sleep loss may require testing in excess of the 25 minutes employed in their study. Bohnen and Gaillard (1994) commented that the 30 minutes of cognitive testing employed in their study may not have been sufficient to detect performance changes as motivation starts to wane.

2.3.1 Sleep Loss Studies

Sleep loss studies seem to have employed relatively longer duration tasks than studies of the effects of physical exertion. All of the sleep loss studies listed in Table 2.1 demonstrated cognitive performance decrements in at least some conditions. Lingenfelser et al. (1994), Jacques et al. (1990), and Horne and Pettit (1985), all using cognitive tests of at least 25 minutes in duration, demonstrated cognitive performance decrements irrespective of the nature of the cognitive task. Bohnen and Gaillard (1994), and May and Kline (1987) demonstrated decrements only on the less demanding cognitive tasks. However, of the five cognitive tests completed by participants in Linde and Bergstrom's (1992) study, only the 40-minute Raven's Progressive Matrices test indicated a performance decrement. The other tasks (a 3-minute Baddeley Logical Reasoning Test, an immediate recall test, and a number-series induction test, each with an estimated duration of less than five minutes, and a test of effortful information-processing of approximately 20 minutes in length) all failed to demonstrate any performance decrement after sleep deprivation.

Ryman et al. (1985), using the 3-minute Baddeley Logical Reasoning Test, found a performance decrement on the less complex actively-worded statements, but not on the more complex passively worded statements. Finally Blagrove et al. (1995) found a performance decrement on all tests after sleep deprivation, but after sleep reduction, only the embedded figures test (16 minutes in duration) indicated a performance decrement. The extended Baddeley Logical Reasoning test (5 or 10 minutes in duration) and the auditory vigilance task (20 minutes in duration) indicated no performance decrement after sleep reduction.

Table 2.1

Cognitive Task Durations in Studies of the Effects of Fatigue on Cognitive Performance

Study	Cognitive Task	Task Duration (mins)
Physical Exertion Studies:		
Davey (1973)	auditory vigilance	2
Féry et al. (1997)	visual perception and short term memory	0.25
Hancock (1986)	visual perception and recognition	3
McNaughten & Gabbard (1993)	arithmetic computation	1.5
Raviv & Low (1990)	scanning	5
Soetens et al. (1992)	visual perception	15
Sparrow & Wright (1993)	intelligence measures	8
Zervas (1990)	intelligence measure	20
Zervas et al. (1991)	design-matching	15*
Sleep Loss Studies:		
Blagrove et al. (1995) [a]	reasoning, auditory vigilance, and embedded figures tests	5 - 20
Bohnen & Gaillard (1994)	concurrent time estimation and vigilance	30
Horne & Pettit (1985)	auditory vigilance	30
Jacques et al. (1990)	in-training examination	240
Linde & Bergstrom (1992) [a]	variety of cognitive tests - per test	3 - 40
Lingenfelser et al. (1994)	battery of medical tasks	25
May & Kline (1987) [b]	variety of cognitive tests - per test	3 - 7**
Ryman et al. (1985) [a,b]	reasoning test	3

Notes: a included Baddeley Logical Reasoning test (Baddeley, 1968)
b included sleep loss and physical exertion variables
* including two 2.5 minute breaks
** total testing time for battery was less than 30 minutes

In addition to the performance effects of cognitive task duration, the preceding discussion also indicates that, in sleep loss research, there may be a difference in performance on cognitive tasks of different complexity. That is, participants tend to show greater cognitive performance decrement on relatively simple tasks than on relatively complex and demanding tasks (e.g. Bohnen & Gaillard, 1994; May & Kline, 1987).

The relative difficulty of the cognitive tasks is difficult to assess. May and Kline (1987) employed a battery of cognitive tests in their study and classified them according to the relative intellectual demand of the task. They considered that tasks involving logical thought and reasoning were more demanding than visual encoding and scanning tasks, an arithmetic task, a memory span task, and tasks requiring the production of novel responses. It could be argued that the division between tasks of greater and lesser cognitive demand depends on the standard against which the tasks are compared. Nevertheless the relative differences could be considered to reflect the amount of cognitive processing required to successfully complete the task. It has been assumed, for the purposes of the comparison of the studies under review, that intelligence, computation, internal monitoring, and logical reasoning tasks are tasks making relatively greater cognitive demand than measures of perception, recognition, and memory.

Bohnen and Gaillard (1994) found that, after one night of sleep deprivation, participants' performance on the more demanding of the two cognitive tasks employed in their study was unaffected, while performance on the less complex task decremented over time. They explained this performance difference as resulting from the higher motivational value of the more cognitively demanding task. May and Kline (1987) also found a similar performance difference using a battery of short cognitive tests. Specifically they found that, in the sleep-deprived condition, performance was not impaired on the tests of logical thought and reasoning, which were among the more cognitively demanding of the test battery.

May and Kline (1987) comment that the effect may be due to the demanding tasks retaining the interest of the participants. Both Tomporowski and Ellis (1986) and Krueger (1989), in their respective reviews of the effects of physical exertion and sustained operations, indicated that motivational factors were an important and influential variable in the study of the effects of fatigue. For example Horne and Pettit (1985) found that participants in their study, who were provided with monetary incentives, were able to maintain baseline performance levels on an auditory vigilance task for up to 36 hours of sleep deprivation. A no-incentive control group demonstrated a steady performance decrement over the same period. However, after 36 hours of deprivation the performance of the group receiving the incentive decremented, until, at 72 hours of deprivation, their performance was no longer distinguishable from that of the control group. That is, motivation may influence performance in the short term, but sleep loss will eventually extinguish the effect.

2.3.2 Physical Exertion Studies

The pattern of long-duration cognitive tests disclosing performance decrements in sleep loss studies is not evident in the physical exertion studies. Given that the longest duration of any of the cognitive tasks in the physical exertion studies listed in Table 2.1 was only 20 minutes, this result may not be surprising.

The three studies with the longest cognitive task durations were Soetens et al. (1992), Zervas (1990), and Zervas et al. (1991). Soetens et al. (1992) found a performance decrement on a relatively less demanding cognitive task, Zervas (1990) found no effect using a demanding task, and Zervas et al. (1991) found a performance improvement on a less demanding task.

The remaining studies, which employed cognitive tasks of eight minutes or less, variously found improvements, decrements, or no effect after physical exertion. Specifically, Davey (1973), Hancock (1986), and Féry et al. (1997) found cognitive performance decrements, McNaughten and Gabbard (1993) found an improvement, and Raviv and Low (1990) and Sparrow and Wright (1993) found no effect. In terms of the relative cognitive demand of the tests employed, McNaughten and Gabbard (1993) and Sparrow and Wright (1993) employed relatively more demanding tasks. The results of neither study indicated that there was a cognitive performance decrement.

It would seem that a relatively long cognitive task may be required to uncover a pattern of performance change due to fatigue.

2.4 Factors Affecting Cognitive Performance

A number of factors have been demonstrated to affect results on tests of cognitive performance, and have the ability to confound the results of a study of the after-effects of exertion on cognitive performance. They are as follows: the type and duration of the physical exertion (and whether it was successful in inducing "fatigue"), the nature, timing and duration of the cognitive performance tasks, non-random individual differences in the participants (such as physical fitness), other types of "fatigue" inducing factors (that is, sleep loss, circadian rhythm disruption, and cognitive exertion), and the concurrent outcomes of fatigue and arousal from exertion. In addition there is the question of whether motivation and effort expended can be controlled or monitored. These factors can be expressed in the form of a model, as shown in Figure 2.1.

The model indicates a possible relationship between fatigue inducing factors and cognitive performance. Mediating factors are fatigue and arousal, the direct outcomes of the fatigue inducing factors. Factors that moderate the relationship and its measurement are the nature of the cognitive measurement method, and the motivation and resulting effort on the part of the participant. Underlying these factors are time elements (including the timing and duration of the fatigue induction procedure, the recovery period, and the duration of the cognitive task), and systematic individual differences between participants (such as physical fitness, age, etc.).

Issues unaddressed by the model in its present form include the exclusion of any bi-directional relationships, or feedback loops, the assumption of the serial nature of the relationships, and the strength and direction of the relationships between individual components of the model. Nevertheless, such a model provides a basis on which the present study can proceed.

The present study is intended to examine the relationship between the after-effects of physical exertion and cognitive performance. In order to investigate this relationship, the following factors, arising from the model, need to be controlled or monitored: consistent use of an appropriate cognitive measure (in terms of complexity, duration, and timing), other fatigue inducing factors (i.e., sleep loss, circadian rhythm disruption, and cognitive exertion), effort expended, individual differences, the physical exertion manipulation, and the recovery period.

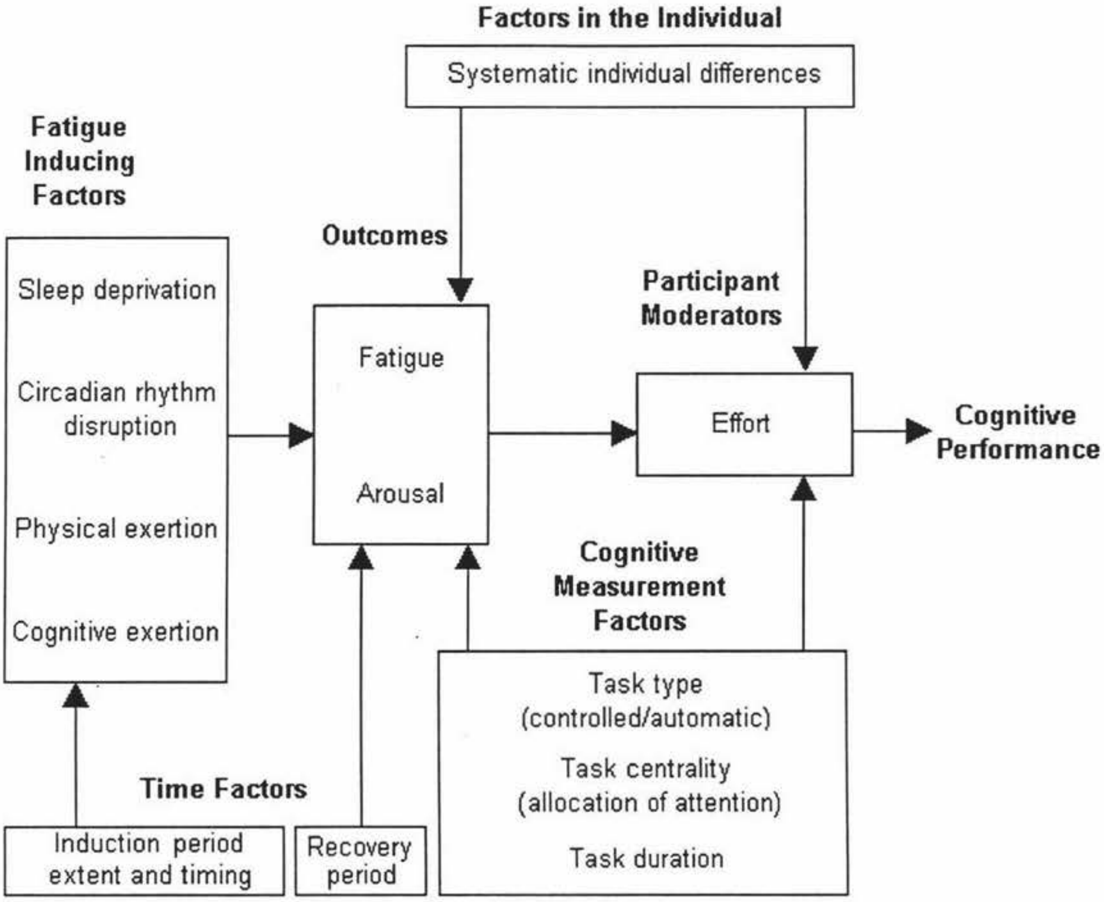

Figure 2.1. Model of relationship between fatigue-inducing factors and cognitive performance.

2.5 Study of the After-Effects of Physical Exertion on Cognitive Performance

2.5.1 Sailboat Racing

Modern sailboat racing is a demanding sport both physically and mentally. Racers can be expected to cope with the physical demands of holding the boat upright and controlling and handling sails in windy conditions. In addition, racers must deal with the mental demands of analysing and predicting wind and sea conditions, and deciding how best to use this information, in combination with the rules and the actions of other boats, to place well in the race. Due to the large variety of sailboat types, the demand can vary considerably from boat type to boat type. Further, both the sailing conditions and the race length can vary considerably. For instance, the demands of racing an 18 metre long boat single-handed around the world are quite different to the demands of sailing a small boat in a Sunday afternoon club race in light conditions.

Nevertheless, sailboat racing can be differentiated from other sporting endeavours in a number of respects. Firstly, the race event is generally of relatively long duration. Races usually require sailors to be on the water for between an hour and three months. Excluding any distance races, the normal time on the water could be expected to be between one and six hours on any given race day. Secondly, the physical demands of racing tend to be a combination of a steady, low-level load, required to maximise the boat's stability, and short bursts of intense demand during rapid changes in course and handling sail adjustments and changes. Thirdly, due to the large volumes of information to be attended too, and the tight time imperatives associated with many of the multiple decisions that have to be made, the cognitive demands of the sport are relatively high.

The participants in the present study were racing sailors who sailed small (6 to 7 metre) keelboats in youth sailing programs. These sailors spend 4 to 5 hours per sailing day on the water sailing drills and races. The sailors completed an extended version of the Baddeley Logical Reasoning test, the Grammatical Transformation Test (see Appendices A and B), before, and after, a day of onshore training and simulated racing, and a day of on-the-water racing and training. The difference between the "before" and "after" test scores on the day ashore could then be compared to the difference scores after a day of sailing. The sailing was expected to provide a task that was moderately physically demanding, and of long duration.

2.5.2 Cognitive Performance Model

Figure 2.1 provided a model of the relationship between fatigue inducing factors and cognitive performance. The present study is intended to investigate one part of this model; the relationship between physical exertion and cognitive performance. To achieve this goal requires that possible confounding elements in the model be controlled, and that the both the independent (physical exertion) and dependent (cognitive performance) variables be satisfactorily operationalised.

The fatigue inducing factors/cognitive performance model in Figure 2.1 discloses a range of factors that can influence cognitive performance, other than physical exertion. The first group of these variables is other fatigue inducing factors. In the present study, sleep loss data were not collected. Nevertheless, sleep loss is not expected to influence results because pre- and post-exertion testing occurred on the same day, with the participants acting as their own control group. In addition, the range and consistency of the days on which testing occurred (four Saturdays and three Sundays) should minimise the chance of a systematic variation in sleep loss. Participants were tested in three groups. Groups were tested on separate days, always either a Saturday or a Sunday, and the day-ashore and sailing days were counter balanced. Pre- and post-exertion testing was completed at approximately the same time each day, which controlled for any circadian effects. Cognitive exertion was controlled in the study by matching, as far as possible, the cognitive tasks of the ashore and sailing days, by the use, on the ashore day, of racing and training simulations, without the physical exertion.

Cognitive measurement factors and task duration factors were controlled through the consistent use of a single test (in this case the extended Baddeley test). Motivation and effort were controlled through two mechanisms. Firstly, a measure of the subjective effort made by the participant in completing each test administration was used to track overall effort. Secondly, the extended length of the test, and the partitioning of the test duration into discrete time periods, was intended reduce motivational masking of fatigue effects by the end of the testing period. The recovery period after exertion was consistent amongst the groups, and the use of a within-subjects design was intended as the primary control mechanism against the confounding effects of individual differences.

Physical exertion was operationalised as approximately four hours of sailboat racing.

2.5.3 Outcome Measures and Hypotheses

In a study of the effects of physical exertion on cognitive performance, it would seem that the most appropriate cognitive test would be one that was of relatively long duration, and that allows the researcher to distinguish between performance on tasks of varying difficulty. The Baddeley Logical Reasoning Test (Baddeley, 1968) was modified to provide such a measure - the 30-minute Grammatical Transformation Test (GTT). The development and pilot of the test is described in Chapter 3. Cognitive performance was operationalised as the difference in performance between administrations of the GTT.

The GTT provides a range of outcome values by which the effects of physical exertion on cognitive performance could be evaluated. Researchers who have used the Baddeley in their fatigue research have used both the number correct (e.g. Linde & Bergstrom, 1992), and the accuracy rate (Ryman et al., 1985) as measures of performance.

In addition to overall results, the data that can be analysed by comparing performance over the duration of the test (with the test divided into four sections, based on elapsed time), and by comparing performance on relatively simple questions to performance on the more complex questions. Further, the results of the tests can be compared to the measures of physical and mental tiredness, and effort expended, provided by the participants. Finally, the results can be examined for evidence of any practice or circadian effects.

It is expected that participants will perform less well on the GTT after a day of racing, than after a day of simulated racing ashore. It is expected performance late in the test will decrease due to an expected reduction in both motivation and arousal, and an increase in cognitive fatigue. It is further expected that when comparing sailing and non-sailing day results, participants will do less well on the relatively simple questions (as compared to relatively complex questions) after the day on the water.

It is expected that there will be evidence of a long-term practice effect that is expected to improve performance over the course of the four tests. It is also expected that there will be an improvement between morning and afternoon test scores due to a short-term practice benefit from the recency of the morning administrations.

3. THE GRAMMATICAL TRANSFORMATION TEST

3.1 Development

The Baddeley Logical Reasoning Test (BLRT) (Baddeley, 1968), on which the GTT is based, consists of a list of statements each of which describes a pair of letters that follow the statement. The respondent is required to identify whether or not the statement is an accurate description of the letter-pair and check the appropriate "True" or "False" column on the answer sheet. An example might be "A follows B - AB". In this case the correct response would be to check "False". The test varies the forms of the sentences by using all possible combinations of the following binary alternatives: (1) affirmative or negative wording, (2) active or passive language, (3) use of either "precedes" or "follows" as the sentence verb, (4) which of the two letters is mentioned first in the sentence, and (5) the order of the letter pair. The sixth binary condition, whether the statement is true or false, is an outcome. That is, it is determined by the combination of the previous five conditions. It should be noted that there has been some disagreement as to the number of possible unique combinations that can be generated by the test. Baddeley (1968) contends that all six conditions can be varied and that there are therefore 64 possible unique combinations. Ryman et al. (1985) generated only 16 combinations, but with variation of the letter order, either within the sentence or in the letter-pair, would have been able to generate 32 unique combinations. It is argued that 32 unique combinations is the maximum number that can be generated by the available binary combinations.

Blagrove et al. (1995) constructed 5 and 10 minute extended versions of the Baddeley (1968) test by introducing a range of variations on the original BLRT statements. Firstly, as well as using the original form of the statements, Blagrove et al. (1995) also employed conditional statements. Original test statements were modified such that a statement may be preceded by the word "if", and an instruction was added to the end of the statement, and prior to the letter pair, directing the respondent to tick the "true" or "false" column, should the condition be true. An example is:

If C precedes M then tick false - CM

Since this statement is true then the respondent should tick false. If the statement is not true then the respondent should mark neither the true nor the false column.

Secondly, Blagrove et al. (1995) made use of two further comparisons of the letters in the pair: whether a letter was to the left of (or to the right of) the other, and whether a letter was larger (or smaller) than the other letter. Examples are:

M is larger than C - mC

C is to the left of M - CM

The effects of introducing the conditional statements were multiple. Firstly, the addition of another variable (the conditional "if" statement) meant that the number of possible unique combinations of variables tripled. That is, in addition to the original 32 possible combinations, the "if" condition allows the respondent to be instructed to mark either the true or the false column if the statement is true, each option providing another 32 combinations. In addition, the two new comparisons ("larger/smaller" and "left of/right of") provided two further complete sets of unique statements with all the variables of the original test (plus the conditional option) available, except for the active/passive wording variable. In addition, the "larger than/smaller than" variable provided the option of varying the order of the latter pair, in addition to varying the comparative size of the letters. In total, these additional comparisons provided another 192 possible combinations. The test now had 288 possible unique combinations.

Secondly, the use of the conditional "if" meant that the number of possible responses to any statement was increased from two (true/false) to three (true/false/no response), and therefore reduced the ability of the respondents to achieve a high score merely by guessing.

Thirdly, the conditional "if" provided a further increase in the level of complexity compared to that available in the original test statements. That is, the most complex statement in the original BLRT was a negatively and passively worded statement. Such a statement could now also be a conditional statement.

Ryman et al. (1985), in their study of the effects of exertion and sleep reduction, employed this distinction between more and less demanding questions. They compared the accuracy of the responses to actively worded and passively worded statements and found a performance decrement after sleep loss of the actively worded, but not the (more complex) passively worded statements. The extended version of the test, by providing an even more complex option, would allow the comparison of performance on simple non-conditional, actively and positively worded statements, versus conditional, passively and negatively worded statements. These questions amounted to approximately one-third of the total number of questions in the test.

A sample of the resulting extended test, the GTT, is included in Appendix A. The sample consists of the covering instruction page of test version one (two versions were used in the present study), and the first 50 statements from the first page of the test.

3.2 Pilot Study

A pilot study was conducted using the GTT and the BLRT. The aims of the study were to evaluate the effectiveness of the GTT in a live environment, to refine the test format as a result of observation and of comments from participants, and to compare the results of the GTT with the BLRT.

3.2.1 Method

Participants

Nine second-year psychology students (1 male and 8 females) from Massey University, Albany volunteered to participate in the pilot study, and completed the first testing session. Six of the nine participants returned to complete the second session of testing. The participants were estimated to be between 20 and 35 years of age.

Materials

The tests employed were an unmodified version of the 3-minute BLRT, and two versions of the 30-minute GTT. The first version of the GTT was modified as a result of the first pilot administration. The modifications involved the reduction of the number of questions from 512 to 378, and the introduction of a marking schedule, for the researcher's use, on the front page of the test, immediately under the instructions. This modified version is shown in Appendix A, and is the form used in the main study.

Procedure

Participants completed an administration of the BLRT followed by the GTT on each of two consecutive Fridays at 1pm. Verbal instructions were provided prior to the administration of each test. Participants then read through the written instructions on the front of each test, and were asked if they had any questions. When all questions had been answered the participants were instructed to start the test. At the end of the allotted time, participants were instructed to stop writing. They were then asked to supply comments as to how the tests or the instructions could be improved.

The second session of test administrations was similar to the first but included changes to the instructions and test version. Specifically, the modified version of the GTT was used, and the verbal instructions had been amended such that participants were requested to supply subjective mental and physical tiredness ratings prior to starting the GTT.

3.2.2 Results and Discussion

There were no major difficulties found in using the test in a pilot study. As a result of observations made in the administration and marking of the tests, the previously discussed modifications were made to the original version of the test. The maximum number of questions attempted by any participant in the first test administration was 277. As a result the test was shortened from 512 questions and 10 pages, to 378 questions and 7 pages. However, in the second administration the maximum number attempted was 365, therefore two additional pages (a total of a further 108 questions) were prepared for any participants in the main study who may require them.

The mean number of questions that were attempted and correctly answered in the first administration of the GTT was 204 and 182 respectively. The comparable figures for the second administration were 238 and 219. This outcome indicated that there may be a practice effect, even though the test administrations were a week apart.

Consequently, those participating in the main study were asked to complete a 10-minute practice version of the GTT at least a week prior to completing the main study tests.

The verbal instructions were modified as a result of comments from the participants, and the inclusion of three further measures on the test. The instructions used in the main study are included in Appendix B. The additional measures provided an assessment of the subjective physical and mental "tiredness" of the participants, and the amount of subjective effort participants put in to completing the test.

Given the very low numbers completing the tests on both days (and particularly on the second day), any conclusions must be tentative. For the six people that completed both administrations there was a strong correlation between the number of correctly answered questions on the first and second tests, $r = .88$. Although this does not provide compelling evidence of the reliability of the GTT, the result provides some assurance in the absence of data from a larger sample.

A comparison of the GTT and the BLRT in respect of the number of questions attempted and the number answered correctly demonstrated moderate to strong correlations between the two tests. On the first testing day (with nine participants completing both tests), the correlations for the number attempted and the number answered correctly were $r = .93$, and $r = .86$ respectively. The comparable values on the second day of testing (with only six participants completing tests) were $r = .54$ and $r = .76$ respectively. These results provide some indication that the differences between the BLRT and the GTT are not so great that comparisons between the tests' results would be meaningless.

4. METHOD

4.1 Participants

Fifty-two sailors (39 males and 13 females) enrolled in youth training programs with the Royal New Zealand Yacht Squadron (RNZYS), and the Bucklands Beach Yacht Club (BBYC), agreed to participate in the study on a voluntary basis. They ranged in age from 16 to 20 years. Thirty-seven sailors completed a full set of four GTTs. Due to a commitment to complete anonymity with regard to test results, the demographic data for those who did not complete all four tests could not be isolated.

Sailors volunteered as participants in response to presentations made to the respective training program groups. The presentations consisted of a description of the research, and discussion of the voluntary basis of participation, the absolute right to withdraw at any time, and the anonymity and confidentiality of the results. Sailors were then invited to participate, and supplied with information sheets and consent forms. All participating sailors prior to their inclusion in the study completed the consent forms. Approximately three-quarters of the sailors involved in the respective training programs agreed to participate in the study.

4.2 Materials

Three versions of the GTT were used in the study. The practice test version contained all the elements of the main tests, except that there were only 108 questions provided for completion, and the duration of the test was 10 minutes. The remaining two versions were identical to each other in all details except the order in which the questions were presented. Two versions of the test were used to minimise the opportunity for memorising answer sequences.

At each administration, in addition to completing the GTT, each participant supplied two uni-dimensional subjective tiredness measures, and a subjective assessment of the effort the participant put in to completing the test. The tiredness measures, one each for physical and mental tiredness, were provided prior to starting the test, while the effort measure was provided at the conclusion of the test. In each case the participant was instructed verbally to assess the value s/he wished to assign to the measure, based on a 1 to 7 scale, and to write this on the front page of the test.

The tiredness measures ranged from fully refreshed (1) to completely exhausted (7) with 4 as the central value. The effort measure was assessed on the basis of comparing effort in the present administration with the effort expended in the practice test. Four equated to a level of effort equal to that put into the practice test, with 1 representing a very large increase in effort, and 7 representing no effort at all. The detailed verbal instructions are included in Appendix B.

The yacht racing simulation program used on the lay-day was Tactics and Strategy Simulator software (Posey Yacht Design, 1992) run on Microsoft Windows 3.1 and Windows 95 based personal computers. The boats used by the sailors on the sailing days were the RNZYS "Elliott 5.9" metre three-person lifting-keel yachts, and the BBYC "J24" five-person keelboats.

All administrations of the GTT, and the activities on the lay-day (ashore) were completed in the clubrooms of the respective yacht clubs.

4.3 Procedure

Participants completed one practice version and four 30-minute administrations of the GTT over three days, each day being separated by at least one week. Participants were tested in three groups: two in mid-1998 (one from each of the two participating yacht clubs), and a third group in mid-1999 (from the RNZYS). Seventeen sailors in the 1998 RNZYS group, and 8 of the 1999 group, completed all four GTT administrations. Twelve sailors from the BBYC group completed the four tests.

The first GTT administration consisted of the shorter practice form of the test, and was administered on a separate occasion to the experimental test administrations. The purpose of the test was to provide a familiarity with the form of the test and the verbal instructions, and to minimise any practice effect prior participants providing data to be used in the main study. The practice form of the GTT was 10 minutes in duration.

One participant who supplied data used in the main study did not complete a practice test.

Participants were then tested, using the 30-minute version of the GTT, in the morning and the afternoons of two weekend days: one day spent on the water (the "sailing day") and one spent on the shore (the "lay-day"). During the administration of the GTT the participants were required to indicate the number of the question they were currently working on at four points in the test. These were at the 3-, 10-, 20-, and 30-minute points in the test, when participants were requested write a number, given by the administrator, in the left-hand margin of the test corresponding to the question that they were working on at the time of the interruption.

The order of the lay and sailing days, and of the two versions of the GTT, was counterbalanced. Twenty sailors undertook the lay-day first and 17 sailors undertook the sailing day first. A sailing day consisted of a normal scheduled day of the respective sailing program, except for the administration of the GTT, which was completed prior to leaving the yacht club to go sailing, and again upon return to the yacht club in the afternoon. An indicative program for the sailing days is described in Table 4.1

Table 4.1

Program for Sailing Days

Time	Activity
9.00am to 9.30am	Arrival at club for briefing for the day's sailing
9.30am to 10.10am	Start time of morning GTT
10.00am to 10.45am	Transfer to dock to prepare boats and go sailing
3.30pm to 4.00pm	Return to dock and pack up boats
4.15pm to 5.00pm	Start time of afternoon GTT

The sailing activities and conditions on each of the three sailing days varied. The BBYC group sailed in up to 15 knots of wind and completed training drills and a harbour course race lasting approximately one hour. Total time on the water was approximately five hours. The 1998 RNZYS group also completed a series of training drills and a shorter Olympic style race and was on the water approximately 4.5 hours in winds of up to 12 knots.

The 1999 RNZYS group completed the sailing day testing over the course of two days. After completing the pre-sailing administration of the GTT, the wind did not strengthen sufficiently to allow the completion of a full sailing day. As a result, the post-sailing test was completed at the end of a subsequent sailing day. That sailing day consisted of a day of drills and short-course racing, and the total time on the water was approximately 4.5 hours. Wind conditions were light with winds speeds of up to 10 knots.

The lay-day program was designed to emulate, as far as possible, the sailing day program, but without the physical exertion. Maintaining the interest of a large group of teenage sailors was also important, both for the accurate simulation of the cognitive element of sailing and in order to maintain commitment. An indicative program for the lay-days is described in Table 4.2.

Table 4:2

Program for Lay-Days

Time	Activity
9.00am to 9.30am	Arrival at club for briefing
9.10am to 9.50am	Start time of morning GTT
9.45am to 10.30am	Commencement of lay-day program
3.10pm to 3.50pm	Start time of afternoon GTT

The lay-day program consisted of coaching activities such as lectures, group exercises, and a videotape presentation, and simulated racing using personal computers and a yacht racing simulation program.

4.4 Data Analysis

The completed tests were marked by comparing the answers provided with the correct answers displayed on transparencies overlaid on each test sheet. The final results for each test were entered into a Microsoft Excel 97 (1996) spreadsheet. The results for each test consisted of the total attempted and total correct for the entire test, for the first three minutes of the test, and for the first and third 10-minute test phases. The score for the second test phase was calculated by deducting the scores from the first and third phases from the total score. The same information was entered for the scores on both simple and complex questions. Finally, the ratings provided for the tiredness and effort measures were entered.

Each set of test scores was hash totalled to provide a check on the accuracy of the data entry. In addition, the (calculated) scores for simple and complex question responses in the second test phase were compared to one-third of the total scores. Any tests with a variance greater than 15% of the total scores were re-marked and the totalling was checked.

The SAS System (1989) computer software (release 6.12) was used for all data screening and analysis.

5. RESULTS

Of the 52 participants who provided data for the study, 37 completed all four administrations of the Grammatical Transformation Test (GTT). All incomplete data sets were excluded from the results. The results provided by a further four participants were excluded as a result of data screening (discussed below) leaving 33 sets of test results for analysis. For a moderate effect size (0.3), the power of the study is estimated at .78 (Howell, 1997). A significance level of $p < .05$ was used for all statistical tests.

Prior to analysis, sets of data resulting from the four administrations of the GTT were investigated for compliance with assumptions of normality. Following discussion of the data screening, the chapter presents the results grouped according to the nature of the measure; that is, the tiredness and effort ratings, the total number of correctly answered questions, and the accuracy of responses for the GTT. Within the investigations of the GTT data, analyses are presented for the full 30-minute GTT, and, for the purposes of comparison with other research, for the first three minutes of the test. Additionally the data are examined for the presence of any practice effect by re-examining it sorted into the chronological order of test presentation. Finally, as part of the accuracy analyses, relative performance on simple and complex questions is evaluated.

5.1 Data Screening

5.1.1 Tiredness and Effort Ratings

Participants supplied subjective evaluations, using a 7-point Likert scale, of their physical and mental tiredness prior to each test administration, and an evaluation of the effort expended in doing the test immediately after its completion. These ratings served as checks on the effectiveness of the manipulation of fatigue state through the sailing and lay-day activities. One participant failed to supply a value for the effort measure in the afternoon administration on the sailing day, and this was estimated.

The probabilities of the physical tiredness, mental tiredness, and effort data having been sampled from a normal population were low (the *p*-values associated with the Wilks-Shapiro statistic, W, ranged from $p = .12$ to $p < .01$, $p = .17$ to $p = .02$, and $p = .03$ to $p < .01$). It is likely that these low values were the result of the narrow range of whole number values selected by the participants. Skewness value ranges for the three measures were -.36 to .31, -.4 to .08, and -.3 to .71. Kurtosis value ranges were -.8 to .29, -.91 to -.02, and -.43 to 1.39. There were no data points that lay outside three inter-quartile ranges from the mean for any of the three subjective measures.

5.1.2 Total Correct Data

In all cases the number of questions correctly answered appeared to have been sampled from a normally distributed population ($p > .05$). Skewness values ranged from .42 to .85, and kurtosis values ranged from –.59 to 1.47. There were no data points that lay outside three inter-quartile ranges from the mean.

The same procedures were applied to the data from the first three minutes of each test administration, producing similar results. The data appeared to have been sampled from a normal population ($p > .05$) in all instances. Likewise there were no data values that fell outside three inter-quartile ranges from the mean. The skewness values ranged from .51 to .87 and the kurtosis values ranged from .04 to .88.

Both the full test (30-minute) data and the 3-minute data were re-sorted into the order in which the tests were administered. These data were to be used to assess any practice effect between days. For the 30-minute data skewness values varied between .24 and.83, while kurtosis values ranged from -.78 to 1.25. Three-minute data skewness values ranged from .38 to 1.29, and kurtosis values ranged from -.29 to 2.61. Both of the high kurtosis values occurred on test occasions that were subject to an outlier value, discussed below.

The data appeared to have been sampled from a normal population ($p > .05$) with the exception of the afternoon of the first test day. In that test, the data for the number of correctly answered questions in the first three minutes had a probability of having been sampled from a normal population of $p = .02$. This may have been the result of an outlying data point that lay between 1.5 and 3 interquartile ranges above the mean. Removal of that data point resulted in a normality probability for the remaining data of $p = .49$. The same participant also supplied a result in the 30-minute test data that was above three inter-quartile ranges from the mean of the practice data in the afternoon of the second day. Removal of the data point altered the normality probability from $p = .06$ to $p = .21$. Both the 30-minute and the 3-minute practice data were tested both with and without the inclusion of the data from this participant (Allison, Gorman, & Primavera, 1993).

5.1.3 Proportion Correct Data

Initial data screening indicated that four of the participants had provided extremely low accuracy rates (number correct/number attempted). Their accuracy across the four test administrations averaged between .41 and .61. All four values were more than 1.5 interquartile ranges below the mean accuracy rate. All other participants had averaged over .75 across the four tests. Accuracy on the tests on the basis of chance was .33. Participants had been verbally instructed prior to each test administration to "concentrate on being as accurate as possible" (see Appendix B). Further, the instructions on the front of each test advised participants that their overall score would be calculated as the total number correct less the total incorrect. The purpose of this statement was to further reinforce the need for accuracy. It is likely that these participants did not understand the instructions, were unable to come to grips with the test, or were careless in their responding. The four test scores were excluded from further data analysis.

The remaining response accuracy rates ($n = 33$) were assessed for normality, and these data were not found to have normal characteristics. No group demonstrated a probability in excess of $p = .01$ that the sampled population was normally distributed. The data groups, in all cases, were negatively skewed (with skewness values ranging from −1.18 to −1.7), with a peaked distribution and outliers on all boxplots of the data.

Due to the likelihood of a ceiling effect, logarithm and inverse transformations (Tabachnick & Fidel, 1989) were performed on the accuracy data. These transformations were unsuccessful in providing data that displayed normal characteristics. The reflected data after a logarithm transformation remained positively skewed, and after an inverse transformation, reverted to a negative skew. In all cases the probability values for a normal distribution remained at less than $p = .01$.

Allison, Gorman, and Primavera (1993) recommend that data that are characterised by violation of parametric assumptions should nevertheless be tested using both parametric and non-parametric tools, and this was the approach that was adopted for accuracy data.

5.2 Tiredness and Effort Ratings Analysis

Participants supplied their subjective assessment of their physical and mental tiredness prior to completing an administration of the GTT and, on completion of the test, the level of effort they expended in doing the test. The data supplied were single measure estimations using a seven point Likert scale with 1 representing maximum effort/minimum tiredness and 7 representing minimum effort/maximum tiredness. Two-way ANOVAs with the day and time of the tests as the independent variables were used to compare participant responses. The means and standard deviations for each of the four tests across each of the three measures are shown in Table 5.1.

Table 5.1

*Means and Standard Deviations for Physical Tiredness, Mental Tiredness, and Test Effort Subjective Measures on a 7-Point Likert Scale**

Subjective Measure	Sailing Day				Lay Day			
	Morning		Afternoon		Morning		Afternoon	
	M	SD	*M*	SD	*M*	SD	*M*	SD
Physical Tiredness	3.46	1.40	4.64	0.95	3.59	1.42	3.61	1.37
Mental Tiredness	3.76	1.48	4.70	1.14	4.09	1.47	4.56	1.42
Test Effort	3.36	1.08	3.26	0.92	3.35	1.36	3.38	1.23

Notes: * 1 represents maximum effort/minimum tiredness and 7 represents minimum effort/maximum tiredness

The comparisons, made using a two-way analysis of variance (ANOVA), revealed that there was no significant difference in effort as a result of the day of the test (*F* [1, 33] = 0.08, $p = .78$), or the time of day (*F* [1, 33] = 0.05, $p = .82$). Nor was there an interaction between the two (*F* [1, 33] = 0.15, $p = .70$). For the physical tiredness measures, however, the main effect of day was significant (*F* [1, 33] = 4.95, $p = .03$) as was the main effect of time of day (*F* [1, 33] = 11.76, $p < .0001$). A significant Day x Time interaction takes precedence over these main effects, *F* [1, 33] = 11.23, $p = .002$. The sailors felt more physically tired on the sailing day (*M* = 4.05, *SD* = 1.0) than on the lay-day (*M* = 3.6, *SD* = 1.19), and in the afternoons (*M* = 4.12, *SD* = 0.91), than in the mornings (*M* = 3.53, *SD* = 1.17. This interaction is graphed in Figure 5.1, which shows that, whereas physical fatigue remained constant across the lay-day, it increased markedly from the morning to the afternoon of the sailing day.

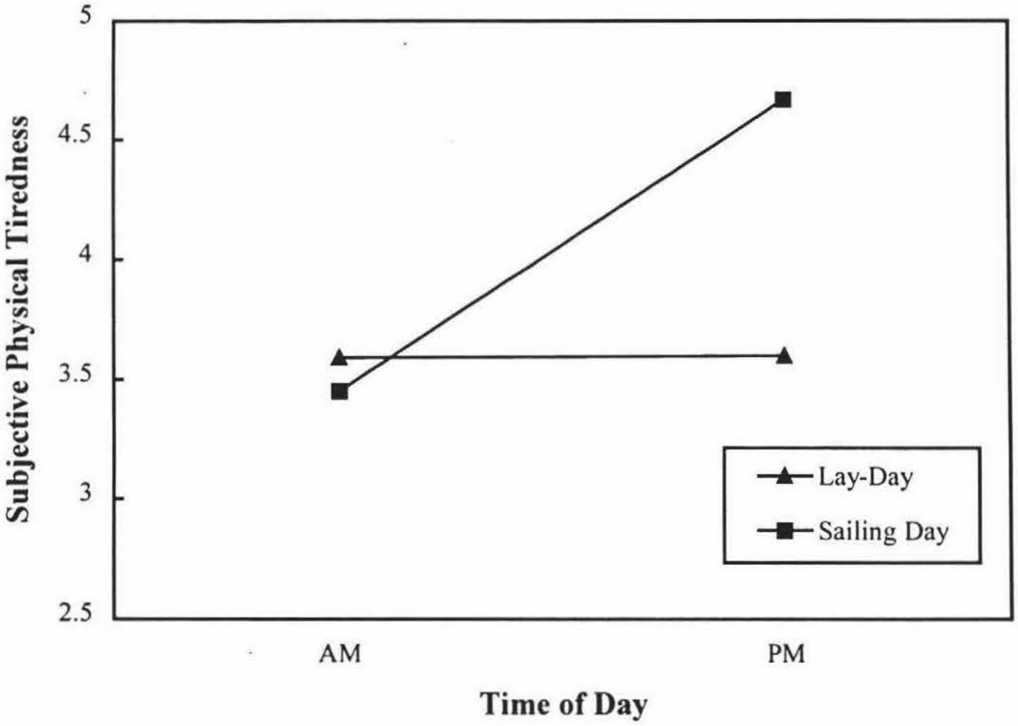

Figure 5.1. Interaction of day and time of test on physical tiredness.

There was also a difference in mental tiredness between the morning and afternoon test administrations, *F* (1, 33) = 10.17, $p = .003$. Sailors felt more mentally tired in the afternoons (*M* = 4.63, *SD* = 1.03) than in the mornings (*M* = 3.92, *SD* =

1.26). Neither the day of the test (*F* [1, 33] = 0.23, $p = .63$) nor the Day x Time interaction (*F* [1, 33] = 1.81, $p = .19$) affected subjective mental tiredness levels.

5.3 Total Correct Data Analysis

5.3.1 Data Sorted According to Test Day Activity

A three-way repeated-measures ANOVA was performed on the number of correct responses for each participant, grouped by phase (0 to 10 minutes, 10 to 20 minutes, and 20 to 30 minutes), day (sailing or lay-day), and time of administration (morning or afternoon). The main effect of the time of administration was significant, *F* (1, 33) = 20.06, $p < .0001$. Sailors correctly answered more questions in the afternoons (*M* = 490.58, *SD* = 162.08) than in the mornings (*M* = 437.82, *SD* = 140.74). There was no significant difference between the number of correctly answered questions on the sailing day (*M* = 470.48, *SD* = 156.81) and the lay-day (*M* = 454.84, *SD* = 173.51), *F* (1, 33) = 0.25, $p = .62$. There was also no significant difference between the first (*M* = 309.09, *SD* = 103.48), second (*M* = 312.61, *SD* = 95.95), and third 10-minute test phases (*M* = 306.7, *SD* = 102.45), *F* (2, 33) = 0.48, $p = .62$.

The number of correctly answered questions was not subject to interactions between day and time (*F* [1, 33] = 0.06, $p = .81$), day and phase (*F* [2, 33] = 2.17, $p = .12$), or day, time and phase (*F* [2, 33] = 2.89, $p = .06$). However, there was a significant Phase x Time interaction, *F* (2, 33) = 4.4, $p = .02$. A graph of this interaction is provided in Figure 5.2. The morning administrations displayed a small increase in the number of correctly answered questions, while the afternoon tests displayed a decrease in correct responses.

A two-way ANOVA, with day and time of administration as the independent variables, was used to analyse the number of questions correctly answered in the first three minutes of each test. The results were similar to those of the full 30-minute test. Specifically, the time of day the test was completed affected the number of questions correctly answered, *F* (1, 33) = 33.63, $p < .0001$. More questions were correctly answered in the afternoon administrations (*M* = 51.2, *SD* = 18.50), than the morning administrations (*M* = 41.58, *SD* = 14.13). However, the number of correctly answered questions was unaffected by whether the sailors were on the water for the day (*M* = 47.52, *SD* = 15.28) or ashore *(M* = 45.26, *SD* = 19.52), *F* (1, 33) = 0.71, *p* = .41. The interaction between day and time was not significant, *F* (1, 33) = 3.49, *p* = .07.

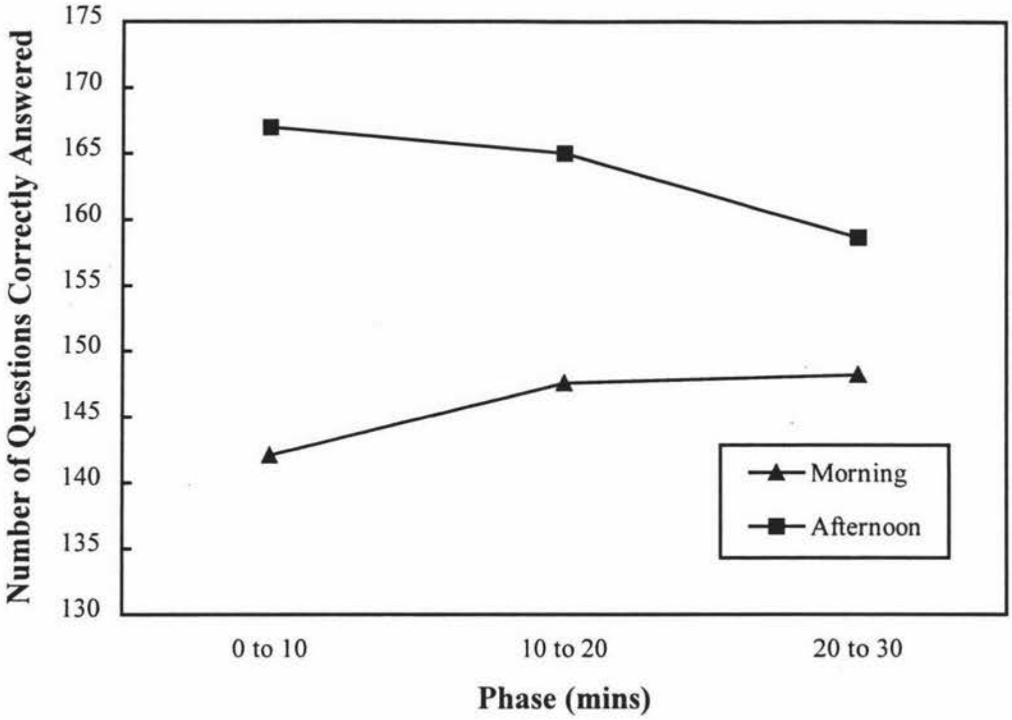

Figure 5.2. Interaction of 10-minute phase and time of test on number of questions correctly answered in the GTT.

5.3.2 Data Sorted According to Order of Test Completion

The influence of the order of test completion was examined. Data were sorted into the order in which the tests were presented; that is, the data were sorted into the first and second days of testing, rather than on the basis of the activities completed on the day. The data were examined using a two-way ANOVA, with chronological test day (first or second, rather than sailing or lay) and the time of testing as the independent variables. Although the main effect of time was not affected by the re-sorting of the data, it was included in the examination to explore any interaction effect with the chronological day the tests occurred.

For the full 30-minute test more questions were correctly answered on the second day of testing (*M* = 492.64, *SD* = 173.51) than on the first day (*M* = 435.76, *SD* = 149.58), *F* (1, 33) = 6.14, *p* = .02. The result was not materially affected by the removal of the single outlier value, *F* (1, 32) = 5.45, *p* = .03. The two-way interaction between sequential day and the time of testing was not significant, *F* (1, 33) = 0.28, *p* = .60.

Analysis of the number of questions correctly answered in the first three minutes of the test also indicated that more questions were correctly answered on the second day of testing (*M* = 50.45, *SD* = 19.26) than on the first (*M* = 42.99, *SD* = 15.85), *F* (1, 33) = 11.23, *p* = .002 Again, the result was not materially affected by the removal of the single outlier value, *F* (1, 32) = 9.77, *p* = .004. However, as illustrated in Figure 5.3, and unlike the full 30-minute test, the 3-minute data demonstrated an interaction effect between the time of day and the day of testing, *F* (1, 33) = 6.38, *p* = .02. Although there was little difference between the mornings of the two days in the number of correctly answered questions, the scores improved more in the afternoon administration on the second test day than in the afternoon of the first testing day.

5.4 Proportion Correct Data Analysis

5.4.1 Statistical Tests

The accuracy of participants in completing the GTT was expressed as a proportion representing the number of questions correctly answered, divided by the number of questions attempted. As previously discussed these data displayed non-normal distribution characteristics. Furthermore, both the accuracy rate and the number of questions attempted can vary between tests (or among the phases within a test). The use of simple averages for the calculation of accuracy rates, without correction for the number of questions attempted may have resulted in inaccurate statistical information.

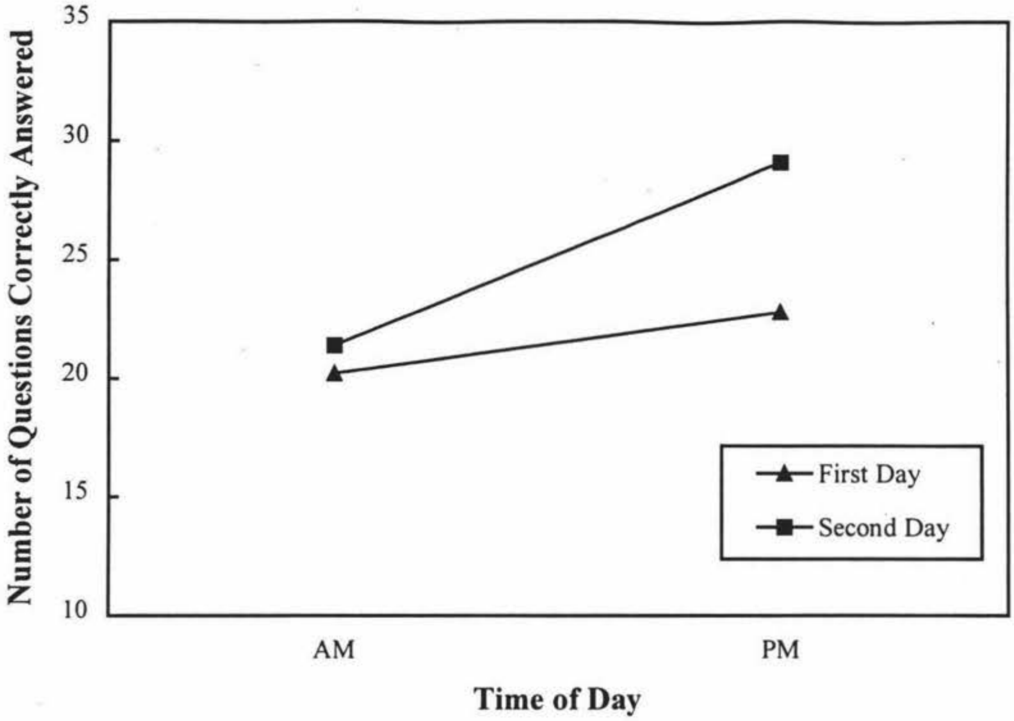

Figure 5.3. Interaction of chronological test day and time of test on number of questions correctly answered in first three minutes of GTT.

Additional statistical testing was completed to ensure there were no material variations in the results arising from the above issues. Two ANOVAs were undertaken for each data set. The first ANOVA included test day and time as the independent variables. A second ANOVA, using, as the independent variable, the four test occasions (as a single variable, in place of day and time), was performed to check for any material distortion of the results due to simple averaging implicit within the ANOVA calculations.

The small differences arose between the results of the two ANOVAs due to this averaging. For example, in the full GTT there was a difference in the (non-significant) F value reported for accuracy as affected by time of testing. Time of testing is reported below as having a value of $F = 0.08$ as part of a three-way ANOVA examining data sorted according to the activity of the day, and as $F = 0.02$ as part of the two-way ANOVA for data sorted into the order of testing. These values would normally be expected to be *identical*. In this instance the difference did not affect the significance or otherwise of the result. This was not universally the case. Although the differences were consistently small, in the case of the main effect of question type, the significance of the result was affected. This is discussed in more detail below.

Additionally, due to the violation of parametric assumptions by the accuracy data, non-parametric tools were used to confirm results for the main effects.

In most cases, the multiple tests confirmed the results of the original ANOVAs. For this reason, and except where test results differ, only the results of the parametric testing, using day and time as the independent variables have been reported. In cases where there have been differences amongst the multiple test outcomes, the results of the alternative tests have also been reported.

5.4.2 Data Sorted According to Test Day Activity

A three-way ANOVA (with day, time, and phase as the independent variables) indicated that there was no overall significant variation in accuracy between the lay-day ($M = .880$, $SD = .065$) and the sailing day ($M = .887$, $SD = .067$), $F\ (1, 33) = 1.19$, $p = .28$. There was also no significant difference in accuracy between the morning ($M = .884$, $SD = .061$) and afternoon administrations ($M = .885$, $SD = .062$), $F\ (1, 33) = 0.08$, $p = .77$. However, accuracy varied amongst the phases, $F\ (2, 33) = 4.74$, $p = .01$. Post-hoc analysis revealed that accuracy in the first 10-minute phase ($M = .872$, $SD = .069$) was lower than in the second ($M = .896$, $SD = .056$) or the third phases ($M = .888$, $SD = .063$). There were no significant two-way interactions between test day and time ($F\ [1, 33] = 1.74$, $p = .20$), day and phase ($F\ [2, 33] = 0.41$, $p = .66$), or time and phase ($F\ [2, 33] = 1.36$, $p = .26$). The interaction between day, time, and phase was also not significant, $F\ (2, 33) = 0.52$, $p = .59$.

The accuracy proportions for the first three minutes of the test indicated a similar pattern to the full 30-minute test results. A two-way ANOVA indicated there was no difference in accuracy between the sailing day ($M = .843$, $SD = .076$) and the lay-day ($M = .841$, $SD = .098$), $F\ (1,33) = 0.17$, $p = .68$. There was also no significant difference between accuracy in the morning ($M = .834$, $SD = .084$) as compared to the afternoon ($M = .848$, $SD = .087$) tests, $F\ (1,33) = 0.48$, $p = .50$. Nor was there any significant interaction between day and time, $F\ (1,33) = 0.77$, $p = .39$.

The accuracy in the first three minutes of the test was compared to the accuracy over the next seven minutes of the test using a two-tailed correlated-samples *t*-test. Accuracy in the first three minutes (*M* = .842, *SD* = .077) was lower than in the next seven minutes (*M* = .887, *SD* = .073), *t* (32) = 4.43, *p* = .0001.

5.4.3 Data Sorted into Order of Test Completion

The proportion correct data were sorted into the order the tests were completed to examine whether the order of test presentation had any effect on accuracy. A two-way ANOVA indicated there was no significant difference between the first (*M* = .876, *SD* = .074) and second (*M* = .891, *SD* = .057) administration days for the 30-minute test, *F* (1,33) = 1.54, *p* = .22. Likewise there was no significant difference in accuracy between the morning (*M* = .884, *SD* = .061) and afternoon administrations (*M* = .885, *SD* = .062), *F* (1,33) = 0.02, *p* = .88. Nor was there any interaction between the day and time of the test, *F* (1,33) = 0.93, *p* = .34.

The same was not, however, true for the three-minute test data. There was no significant difference between the first (*M* = .833, *SD* = .076) and second testing days (*M* = .851, *SD* = .096), *F* (1,33) = 1.30, *p* = .26, or the morning (*M* = .835, *SD* = .084) and afternoon administrations (*M* = .848, *SD* = .087), *F* (1,33) = 0.48, *p* = .50. There was, however, an interaction between the time and the day, *F* (1,33) = 6.03, *p* = .02. This is illustrated in Figure 5.4. Although accuracy was higher on the morning of the first testing day than on the second day, the positions were reversed in the afternoon tests. That is, accuracy declined between the morning and afternoon tests on the first day, but improved on the second.

5.4.4 Simple and Complex Question Data

Comparison of performance on the simpler versus the more complex question types was evaluated only in terms of the accuracy of the responses. The "simple" and "complex" questions were randomly distributed within the tests, and participants completed only a portion of each test. The respective numbers of simple and complex questions attempted would not have been equivalent, and were therefore not evaluated. Further, the data for these analyses consisted of only those questions identified as "simple" or "complex", under the criteria discussed in Chapter 3. This amounted to approximately one-third of the total number of questions answered. Given that the data have been selected for inclusion on the basis of question type, only those main effects and interactions that included the type of question as a variable were considered. All other main effects and interactions have already been assessed with the data from the full test, and have therefore not been re-evaluated.

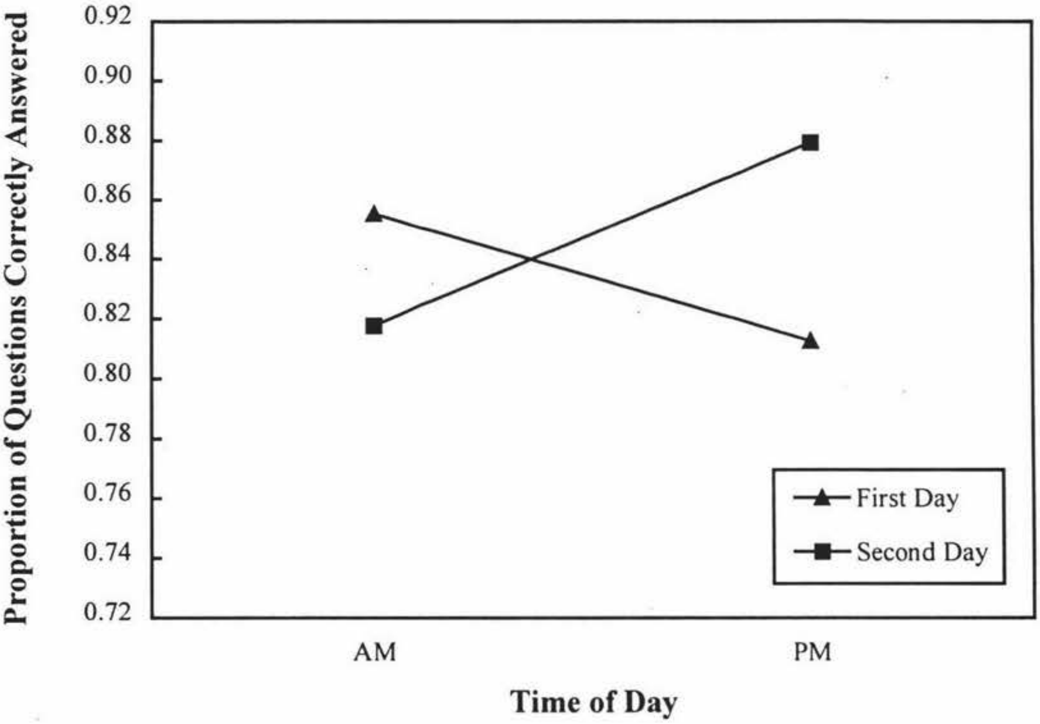

Figure 5.4. Interaction of chronological test day and time of test on response accuracy proportions in first three minutes of GTT.

5.4.4.1 Thirty-Minute Test Data

A four-way ANOVA was used to explore the effects of question type, and the test day, time, and phase on response accuracy. No significant difference in response accuracy was found between the simple and complex question types, F (1, 33) = 3.80, p = .06. However, this was different to the result obtained from a simpler three-way ANOVA using question type, and the test day and time as the independent variables. This second test indicated that simple questions (M = .829, SD = .047) were answered less accurately than complex questions (*M* = .875, SD = .127), *F* (1, 33) = 4.61, *p* = .04. Complex questions also displayed greater variability in response accuracy. A Wilcoxon's Signed Ranks test comparing overall accuracy on simple and complex questions also indicated that simple questions (*Mdn* = .847) were answered less accurately than complex questions (*Mdn* = .918), *T* (32), 186.50, *p* < .0003. The weight of evidence supports the conclusion that accuracy did vary as a result of question type.

Results from the four-way ANOVA indicated that there was no significant two-way interaction between the question type and the test day (*F* [1, 33] = 0.3, *p* = .59). Nor were there any significant three-way interactions between question type and the test day and 10-minute phase (*F* [2, 33] = 0.02, *p* = .98), the time of day and phase (*F* [2, 33] = 0.15, *p* = .86), or the test day and time (*F* [1, 33] = 1.29, *p* = .26).

There were, however, three interactions that reached significance. Firstly there was an interaction between question type and the time of day, *F* (1, 33) = 4.30, *p* = .05. Figure 5.5 illustrates that, whereas accuracy on complex questions improved between the morning and afternoon administrations, accuracy on simple questions declined slightly.

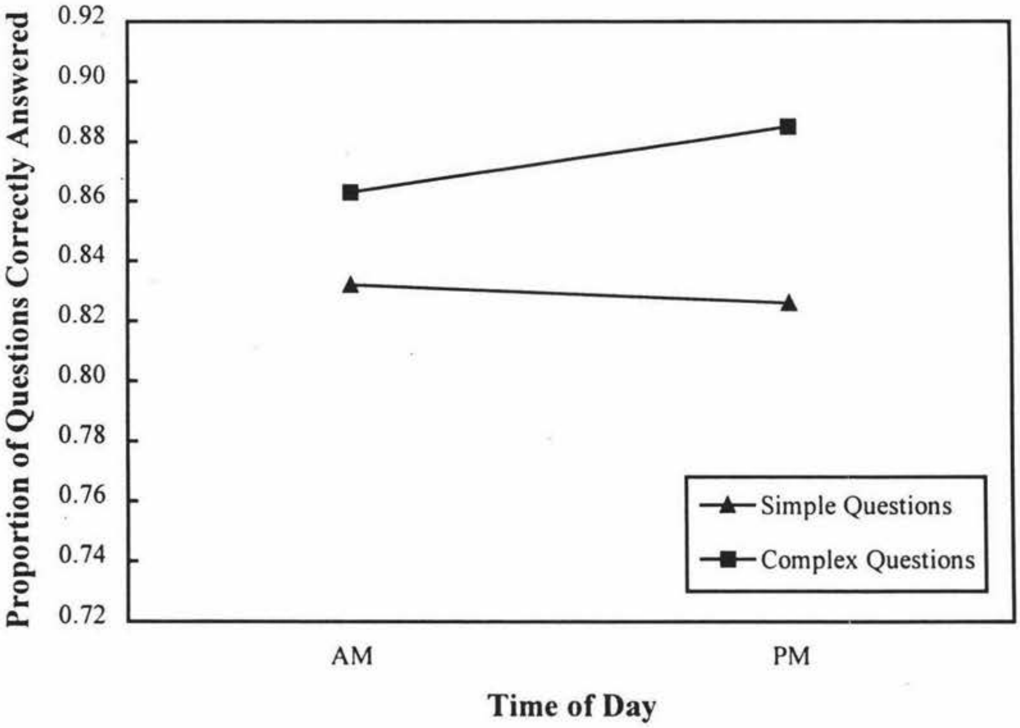

Figure 5.5. Interaction of question type and time of test on response accuracy proportions in the GTT.

Secondly, there was an interaction between question type and phase, *F* (2, 33) = 4.64, *p* = .01. Figure 5.6 illustrates that accuracy on simple questions improved more than on complex questions between the first and second test phases. Relative accuracy for simple and complex questions remained unchanged between the 20- and 30-minute test phases.

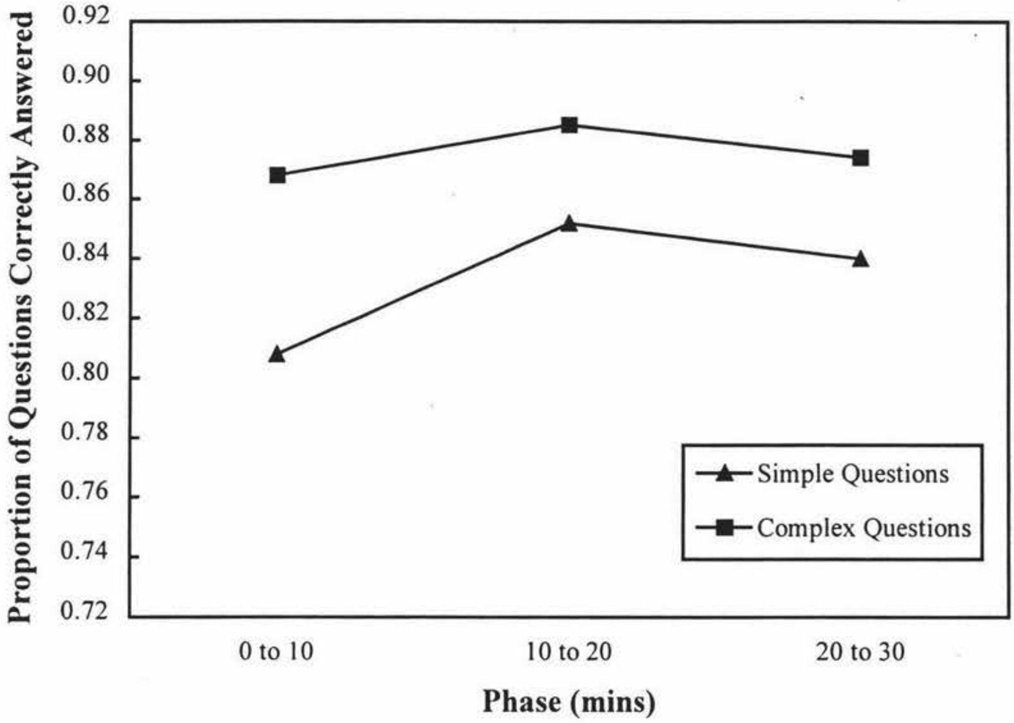

Figure 5.6. Interaction of question type and 10-minute test phase on response accuracy proportions in the GTT.

Finally, there was an interaction between all four independent variables; question type, test day, time, and phase, *F* (2, 33) = 5.45, *p* = .01. The interaction is displayed in Figures 5.7 to 5.10. Accuracy on complex questions stayed comparatively stable over the course of the four tests and across each of the three 10-minute phases within the tests. Likewise, the accuracy for simple questions on the afternoon of the sailing day and the morning of the lay-day, remained relatively stable across the test phases. In contrast, the accuracy of responses to simple questions varied considerably between phases on the morning of the sailing day, and the afternoon of the lay-day. In these latter tests accuracy improved dramatically between the first and second phases and then fell away somewhat in the third phase.

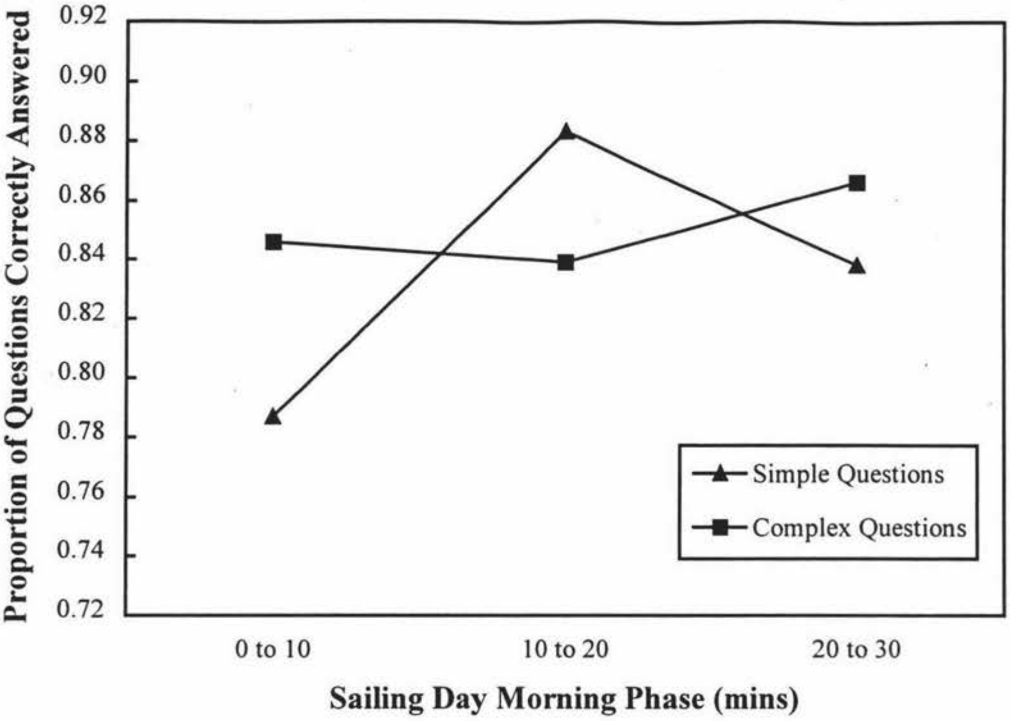

Figure 5.7. Interaction of question type and phase on response accuracy proportions in the GTT on the sailing day morning administration.

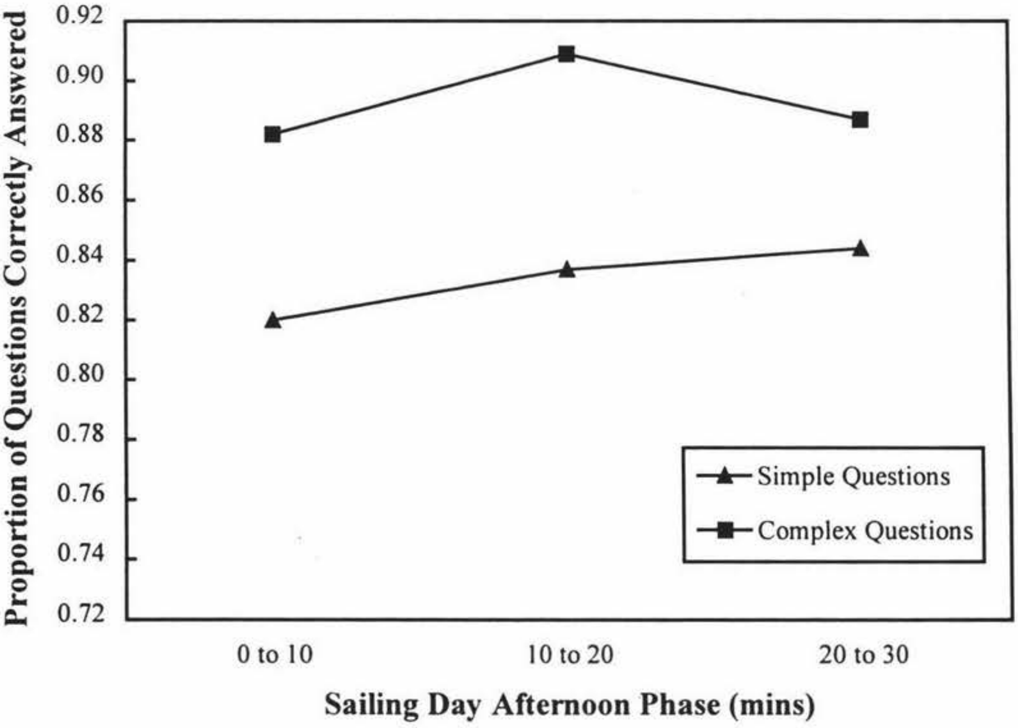

Figure 5.8. Interaction of question type and phase on response accuracy proportions in the GTT on the sailing day afternoon administration.

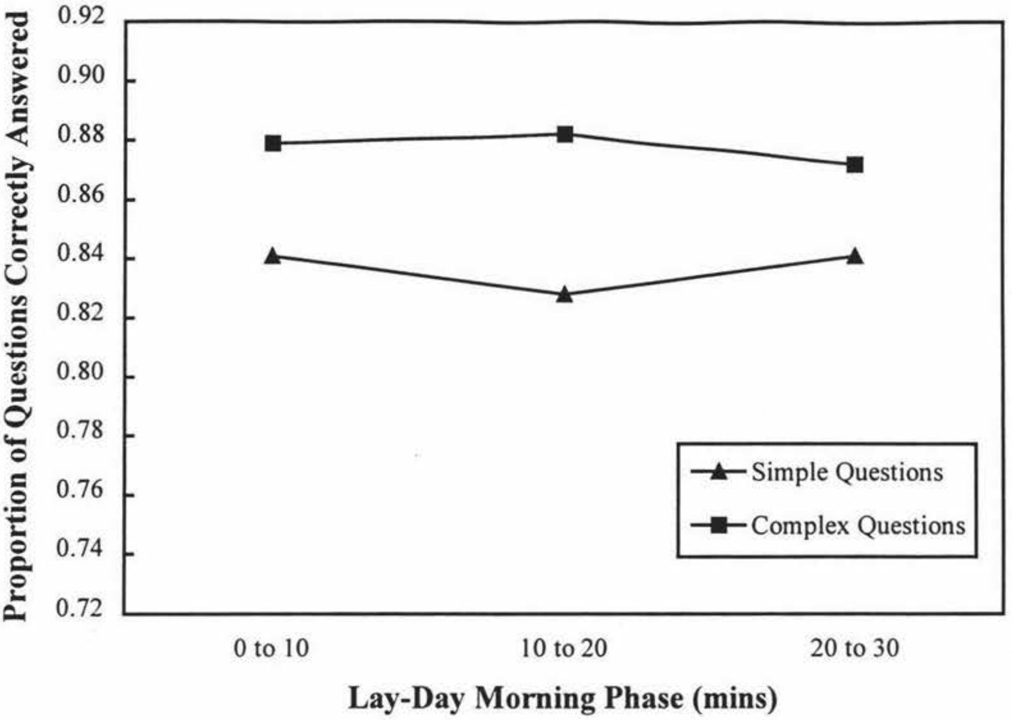

Figure 5.9. Interaction of question type and phase on response accuracy proportions in the GTT on the lay-day morning administration.

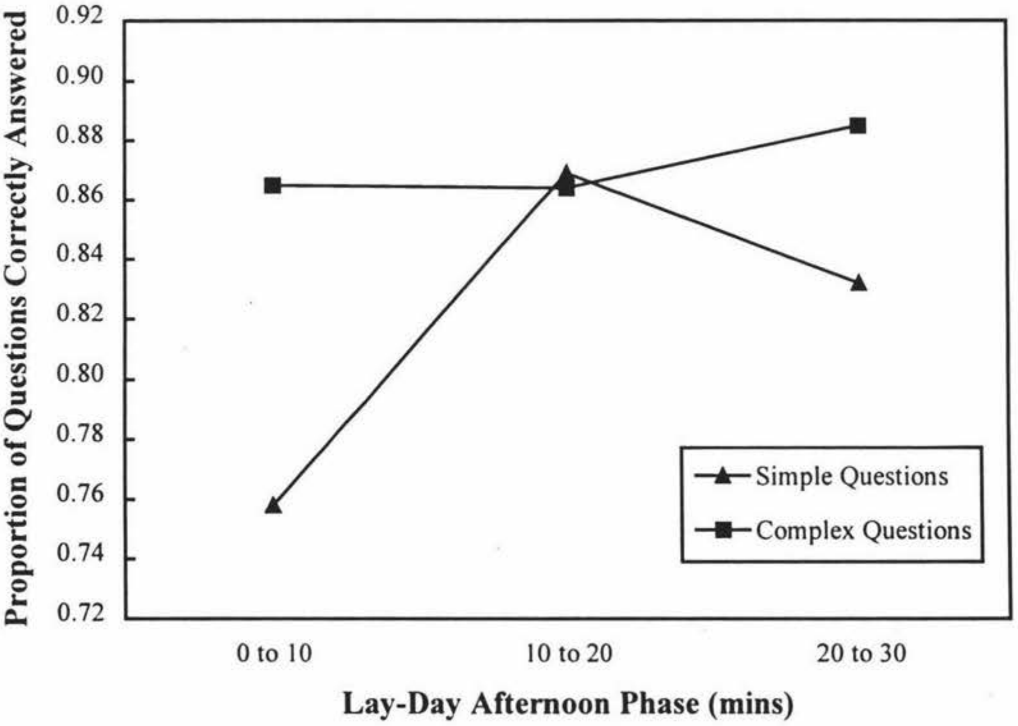

Figure 5.10. Interaction of question type and phase on response accuracy proportions in the GTT on the lay-day afternoon administration.

5.4.4.2 Three-Minute Test Data

A three-way ANOVA was performed to explore the effects of the test day and time on accuracy in answering simple and complex questions. Once again, due to the reduced data set, only those effects that included question type as a variable were considered. There was a significant main effect for question type, F (1,33) = 7.68, p = .01). Simple questions (M = .802, SD = .082) were answered less accurately than complex questions (M = .863, SD = .157). Similar to the 30-minute data, there was greater variability in the accuracy on complex questions than on simple questions. No two-way or three-way interaction effects that included question type were significant. The results for the interactions were question type and test day, F (1,33) = 0, p = .97, question type and test time, F (1,33) = 0.47, p = .50; and question type and test day and time, F (1,33) = 0.06, p = .80.

6. DISCUSSION

It was expected that, using an extended cognitive task (the GTT), it would be possible to demonstrate changes in cognitive performance resulting from the after-effects of physical exertion. These effects could be measured as either changes in the accuracy of responding, or, with the exception of responses to simple versus complex questions, in the total number of correctly answered questions.

Specifically, it was hypothesised that performance on the GTT would be poorer after a day of sailboat racing than after a day of simulated racing ashore, and that performance would be poorer late in the test as compared to early in the test. It was further expected that, when comparing sailing and non-sailing day results, participants would perform less well on the relatively simple questions (as compared to relatively complex questions) after the day on the water. Finally, it was expected that there would be evidence of a practice effect (and performance would improve over time), and a circadian effect (that is, there would be a relationship between test performance and the timing of the administration - morning or afternoon).

Although there was evidence of circadian and practice effects, and a performance difference based on question complexity, performance on the GTT was not influenced by the after-effects of physical exertion.

6.1 Overall Performance

6.1.1 Manipulation Checks and Controls

Of primary importance was evidence that a day of sailboat racing resulted in greater physical tiredness than a day spent ashore in simulated racing conditions. The subjective measure of physical tiredness indicated that the participants did indeed feel more tired after a day on the water. Furthermore, there was no significant difference between the physical tiredness measurements taken on the mornings of the test days. While this does not constitute exhaustive reliability testing of the measure, it would have been cause for some concern had the morning measures differed. However, the result does indicate that, in the morning, perceived tiredness did not seem to be affected by expectations related to the activities the sailors were to undertake during the day ahead.

Additionally, the result of the afternoon physical tiredness measure on the lay-day was not different to the two morning measures. This may indicate that the day ashore did not result in any physical tiredness arising from the activities. Alternatively it may indicate that there may have been unmeasured effects that compensated for any physical tiredness effect of the ashore-based activities. It is assumed, however, that such compensatory effects would have been present on both the sailing and non-sailing days, and the differences between the two afternoon estimates of physical tiredness were a result of a day on the water.

Two other factors that could have directly affected GTT test results were the levels of mental tiredness of the sailors, and any variation in the effort they expended in completing each of the tests. In spite of variations in subjective mental and physical tiredness, the sailors indicated that they had expended equivalent effort in completing each of the four the tests. This result provides some assurance that results were not confounded by variations in effort, rather than variations in physical tiredness.

The pattern for the estimates of mental tiredness also indicates that this variable was not influential on test results. That is, there was no significant difference between the two morning estimates or the two afternoon estimates. However, there was a non-significant difference between the two morning estimates amounting to z = .22. Although this effect size is small (Howell, 1997), it may be worthy of some discussion. The expected difference was nil, and the power of the study to detect an effect of this size was .57. Andersen and Stoove (1998) argue the case for not dismissing non-significant results solely in the basis of p being greater than .05. The power of the study to detect an effect and the context of the result should also be evaluated.

The morning mental tiredness values may have been influenced by a day of the week effect. Due to difficulties with weather conditions, and available days for sailing within the respective youth training programs, counterbalancing did not fully control for differences that may have arisen from using different weekdays. The bulk of the participants (25 of the 37) completed their sailing day tests on a Saturday, while the majority of the sailors (29 of the 37) completed their lay-day tests on a Sunday. It is possible that this group of youth sailors may have found Saturday nights more conducive to partying than Friday nights, and this may have influenced their feelings of mental tiredness early on a Sunday morning; the day on which the majority of the sailors completed their lay-day tests. Should such an influence exist, its effects did not seem to extend to the effort and physical tiredness measures.

The manipulation checks indicate that, when compared to a day of simulated racing ashore, a day of sailing successfully increased physical tiredness without affecting the comparative effort or mental fatigue variables. Attention can now be turned to a comparison of cognitive performance in the post-activity conditions, the principal area of interest.

6.1.2 Effects of Fatigue on Performance

Previous research has not shown that the after-effects of sustained physical exertion result in a predictable pattern of cognitive performance outcomes (Tomporowski & Ellis, 1986). Researchers have found both performance improvements (McNaughton & Gabbard, 1993), and no effect (Raviv & Lowe, 1993; Zervas, 1990) from studies of the effects of long-duration, moderate-intensity, physical exertion tasks. When more strenuous exercise has been used in a long-duration task cognitive performance decrements have been found (Hancock, 1986; Soetens, et al., 1992). It was proposed that inconsistencies in the results of such research might have been due to the length and the complexity of the cognitive task being used.

It was expected that a long-duration repetitive cognitive task (the GTT) would overcome the ability of an individual, by virtue of interest or the employment of additional effort, to compensate for the effects of fatigue. Performance on the GTT was not influenced by the after-effects of physical exertion from sailing; either in the number of correct responses, or response accuracy. This, in spite of the fact that the sailors indicated they were significantly more physically tired after sailing, than after the day ashore.

It appears that the additional length of the GTT, as compared to the shorter cognitive tasks used in previous studies, did not locate any previously undisclosed cognitive performance decrement. Support from the present study for previous research findings was extended by results from the first three minutes of the GTT, which indicated the same pattern of results; that the after-effects of extended moderate physical exertion does not negatively affect cognitive performance.

It was also expected that any detrimental performance effect would tend to be more pronounced nearer the end of the GTT than at the beginning, and that such an effect would interact with the level of physical tiredness of the participants. The GTT was divided into three 10-minute phases to provide a means for making such comparisons. Again, there was no evidence of a cognitive performance decrement in either the total number of questions correctly answered, or in the accuracy of the responses. Analysis of the test performance by phase did, however, reveal two other effects. There was a progressive increase in response accuracy through the early part of the test, and an interaction between the test phase and the time of day in respect of the total number of correct answers. Both of these effects are considered later as part of the discussion of practice and circadian effects.

Tomporowski and Ellis (1986) proposed two explanations for the discrepant findings in their review of the effects of exercise on cognition. Firstly, they proposed that arousal of the central nervous system and physical fatigue are the concurrent results of physical exertion, and that these factors act independently on cognitive performance. Secondly, they proposed that motivational factors affect test performance, and that "exercise per se does not alter cognitive functioning" (p. 344).

The present study attempted to control for motivational factors by the use of a long, boring, and demanding cognitive test. Further, by taking measurements at the three-minute point in the GTT it was hoped that a variation in performance over the length of the test could be demonstrated. There was neither a variation in test performance as a result of fatigue, nor a difference between the full GTT and the 3-minute test results. Although the test results tend to support the proposition that moderate physical fatigue does not influence cognitive performance, the present study offers little in support of the motivational explanation either. Given that motivation over the course of the test was expected to vary, results would also have been expected to change. In conflict with the expected performance decrement, the only performance change over the course of the test was a progressive improvement in accuracy. However, in the present study there was no direct measurement of motivation, and it would be, at best, premature to argue that motivation does not influence test performance.

In relation to the former explanation, that arousal and fatigue act independently on cognitive performance, it has yet to be demonstrated that physical fatigue through long-duration, moderate exertion has any decremental effect on cognitive performance. It may be that negative cognitive effects of physical fatigue occur only in cases of extreme exertion (e.g. Hancock, 1986; Soetens, et al., 1992), but that it is the effects of arousal that influence cognitive test performance after moderate exertion. If this explanation is correct, then, given that the present study attempted to minimise the effects of arousal, it is not surprising that no effect on performance was found.

6.2 Simple Versus Complex Question Responses

Research on the effects of sleep loss on cognitive performance has indicated that fatigue may have different effects on performance depending on the relative difficulty of the cognitive task. Bohnen and Gaillard (1994), May and Kline (1987), and Ryman et al. (1985) all found a greater performance decrement after sleep loss on relatively simpler cognitive tasks than on more complex tasks. It appears that the comparison of relative performance on cognitive tasks of differing complexity has not previously been made in research on the effects of physical fatigue. Ryman et al. (1985) used a variation on the Baddeley Logical Reasoning Test (Baddeley, 1968) and compared response accuracy to relatively simple statements within the test to the relatively complex statements. It was this approach that was employed in the present study.

Overall, relatively simple questions were answered less accurately than relatively complex questions, and the effect size ($z = .52$) was moderate (Howell, 1997). This counter-intuitive effect occurred irrespective of the fatigue condition (the interaction of day and time). The difference also occurred in the first three-minutes of the test. This result differs from the results of the Ryman et al. (1985) study, that used the 3-minute Baddeley as the cognitive measure. Ryman et al. found that comprehension of the simple sentences was better than comprehension of the more complex sentences, but that after sleep loss, the difference between the two was reduced. Specifically, the accuracy of responses to simple questions declined, while accuracy on complex questions was not affected. However, in keeping with the results of the present study, Ryman et al. (1985) found no change in the comprehension of either sentence type after long and demanding physical exertion.

Adequate explanation for poorer performance on simpler questions is difficult. This finding indicates an absolute difference in accuracy rather than a difference in the *change* in performance found in previous sleep loss research (e.g. Bohnen & Gaillard, 1994; May & Kline, 1987). It can only be speculated that a combination of the length and the time pressure of the test may have influenced the relative care taken in answering apparently simpler questions. Given that the sailors were aware of the length of the test, this may also have affected their approach in the first three minutes.

The difference in performance related to question type was affected by interactions with the time of the test, and phase. The interaction with time indicates that complex questions benefit from any time-of-day effect, whereas accuracy on simple questions remains largely unaffected. The interaction between question type and phase indicates that the progressive improvement in accuracy in the early part of the test demonstrated in the overall results, benefits the simple question accuracy more than accuracy on the complex questions.

In addition there was a complex interaction between question type, and the day, time, and phase of the test. The pattern of the interaction is, again difficult to explain adequately. It could have been expected that the pattern of accuracy in the two morning test administrations would have been similar and there may have been a difference between the afternoon tests. However, the similarities that occurred were between the morning of the sailing day and the afternoon of the lay-day, and the morning of the lay-day and the afternoon of the sailing day.

The largest difference between simple and complex question response accuracy occurred on the afternoon of the sailing day. Although this may appear to indicate an effect from sailing activities, there was no significant interaction between question type, day, and time. That is, the significant interaction occurred between question type, day, time, *and* phase. The similarity between the sailing day afternoon and the lay-day morning results further weakens the case for an effect from sailing activities.

During the second test phase of both the sailing day morning and the lay-day afternoon, accuracy on the simple statements was uncharacteristically high, and higher than the accuracy on the complex questions. It is likely that this feature of the data has resulted in the significant four-way interaction, as it is the only feature that appears to include concurrent variation on all four variables; day, time, phase, and question type. As the two versions of the test were counterbalanced within the participant group and as between the two days of activity, the test version unlikely to have influenced the results. Why accuracy on simple questions should be uncharacteristically high in the second phase of only the sailing day morning and the lay-day afternoon administrations is difficult to understand.

In summary, simple questions were answered consistently less accurately than complex questions. In addition, the after-effects of a day of sailing had no effect on the relative accuracy of answers to simple versus complex questions.

6.3 Practice and Time-of-Testing Effects

There was evidence of two effects that could be attributed to practice or learning; an improvement in response accuracy during the course of each test, and, compared to the first testing day, an increase in the total number of questions correctly answered on the second day.

The change in accuracy in the early part of the tests was supported by a progressive improvement from the first 3 minutes to the next 7 minutes. It appears that, irrespective of whether or not the test has been completed previously, there is a "warming up" period with each test. That is, over the course of the first 10 minutes of the test, the sailors steadily improved the accuracy of their responses. After that time accuracy reached a plateau. There was evidence of a practice effect in addition to the main effect of phase. When the data were re-sorted into the order in which the tests were administered, there was an increase in the number of correctly answered questions between the first and second testing days. There was not, however, any change in accuracy between the two days.

Taken together with the phase information, it appears that there was a short-term learning effect for accuracy and a long-term learning effect for the number of correctly answered questions. That is, participants learned how to answer more rapidly after previous exposure to the test but such exposure did not effect accuracy. However, as participants proceeded through each administration, their accuracy improved initially then reached a plateau, while their rate of answering did not change. This improvement in accuracy was a temporary improvement, and did not carry over from one test to the next, even when that next test may have been only a few hours later.

One of the distinctions drawn by Shiffren and Schneider (1977), between controlled processes and automatic processes, was that, as compared to automatic processes, tasks using controlled processes demonstrate an early plateau in performance improvement. It is likely that many tasks represent a mixture of controlled and automatic processing. In the present study there was evidence of both an early improvement plateau (in response accuracy), and an extended and steady improvement in the total number of questions answered over the two testing days. This may serve to indicate that the controlled and automatic components of the test may affect different performance measures. It may also mean that the Baddeley-length test and the GTT are not entirely comparable due to the lack of a short-term plateau in accuracy in the shorter test.

In the design of the present study, counterbalancing of the lay- and sailing-days, and the use of a practice test should have controlled for any long term learning effect in the main results. However, the short-term learning effect for accuracy was not controlled out. Instead it was made visible by the use of phases in the tests. Future studies could eliminate the effect by the use of a brief warm-up test prior to the main test at each administration.

There was also a main effect of the time of the testing (morning or afternoon) for the number of correctly answered questions. Participants correctly answered more questions in the afternoon administrations of the GTT. This effect was elaborated by the interaction of time and phase, which indicated that the gap in performance between morning and afternoon tests narrowed as the tests progress. The cause of this difference cannot be readily isolated. It may represent further evidence of a longer-term practice or learning effect (where learning has a diminishing effect over time). That is, the increase in the number of questions answered correctly from the morning to the afternoon administrations did not differ as between the first and second testing days. This suggests that any learning component in the time-of-day effect is temporary, as the performance increase was the same on both the first and the second testing days.

The difference may also have been the result of circadian changes in such things as physiological arousal or recovery from a hangover. It was not a result of arousal from the physical exertion of sailing as the effect occurred independently of the testing day. Future research could further isolate the circadian effects by counterbalancing the order of the morning and afternoon administrations. The timing of the test did not effect accuracy.

6.4 Comparison with the Baddeley Test

Previous research has used the BLRT to investigate differences in cognitive performance resulting from exertion or sleep loss. The GTT differs from the BLRT in two ways: the length of the test (30 minutes compared to 3 minutes for the BLRT), and the range of question types used in the test (the GTT uses a wider range of questions – see Chapter 3). Participants, during the completion of the GTT, indicated the number of the question they were working on at the three-minute point in the test. This allowed the comparison of the three-minute results with results from the full test and with the results from previous research. The greater range of question types used in the GTT, and the fact that participants knew they had another 27 minutes of testing to complete after the first three minutes of the GTT, necessarily limits the extent to which conclusions can be drawn from comparisons with the shorter test.

There were two differences evident from a comparison of the results of the full 30-minute GTT with the 3-minute results. The first difference was in the evidence for a practice effect using the data sorted into the order of test administration. Similar to the 30-minute data, accuracy in the first three minutes of the test did not vary as a result of different test days or test times. However, unlike the 30-minute results, there was an interaction between day and time in the 3-minute results. On the first test day accuracy declined between the morning and afternoon administrations, while on the second test day accuracy was low in the morning, and improved in the afternoon.

Again, it is difficult to speculate why this may have occurred. Whatever the reason for the variation, it occurred only at the very beginning of the tests, as the effect was not present in the 30-minute test result. It is possible that expectations on the first test day were high, and there may have been an initial enthusiasm that faded rapidly. On the second day, having already completed a day of tests some time previously, the expectations may have been reversed. The enthusiasm in the last test may have been the result of a poor effort in the morning, and a delight in the afternoon test being the final test. It is possible that this effect may not have occurred had the GTT been only three minutes long.

Secondly, unlike the results from the full GTT, there was no interaction between the question type and the time of day at the three-minute point in the test. That is, the time of day the test was completed affected simple and complex questions equally. This indicates that the improvement in complex question accuracy in the afternoons that was found in the full GTT results must have occurred after the first three-minute period.

In comparing these results to Ryman et al. (1985) there is an important difference in the nature of the GTT compared to the BLRT, as regards simple versus complex questions. Ryman et al. (1985) found a difference in the accuracy of participants after sleep loss using the standard 3-minute BLRT. The present study used only one-third of the total number of questions answered for the purposes of comparing simple and complex question accuracy; that is, only those questions that could be clearly identified as relatively simple or complex. The result is that there was only one-third the number of questions available for analysis at the three-minute point in the GTT compared to the Ryman et al. (1985) study. In many cases this resulted in fewer than three examples of a question type being attempted by an individual in the first three minutes. This may have had the effect of reducing the probability of finding differences in the present study by reducing the sensitivity of the measure.

Overall, the 3-minute and the 30-minute test results were similar. The principal difference between the full GTT and the BLRT lies in the effects of prior exposure to the test. The GTT is long enough to expose a plateau in the rising accuracy of participants as they progress through the test. It is possible that this may affect any given individual's comparative score between the BLRT and the GTT depending on how quickly they reach that plateau.

6.5 Model of Fatigue/Cognitive Performance

Figure 2.1 illustrated a proposed model of the relationship between fatigue-inducing factors and cognitive performance. In relation to the after-effects of physical exertion on cognitive performance, a number of factors were expected to mediate or moderate the effect. They were: the type and duration of the physical exertion (and whether it was successful in inducing "fatigue"), the nature, timing and duration of the cognitive performance tasks, non-random individual differences in the participants (such as physical fitness), the amount of motivation and effort expended in relation to the cognitive task, other types of "fatigue" inducing factors (that is, sleep loss, circadian rhythm disruption, and cognitive exertion), and the concurrent outcomes from exertion of fatigue and arousal.

The type and duration of the physical exertion (a day of racing) was expected to induce a fatigue effect. It is in the nature of sailboat racing that physical exertion is moderate for long periods, combined with short bursts of heavy exertion. The time elapsed from finishing the racing to sitting the afternoon test was also expected to minimise the arousal effect. The subjective measure of physical tiredness indicated that the day on the water did provide a greater level of physical tiredness than the day ashore. The mental tiredness measure provided indirect evidence of no significant difference between the level of arousal post-sailing and post-non-sailing. Likewise, the mental tiredness measure provided some assurance that the effect of cognitive exertion was similar as between the sailing and lay-days. However, there was no direct measure of the level of arousal of the participants at the time of the test.

The within-subjects design of the study controlled for the effects of any non-random individual differences. The counterbalancing of the days ashore and on the water should have controlled for any practice effect in the results. Likewise, the completion of tests at the same times on both lay and sailing days would have controlled for any circadian effects. However, one group of eight participants were unable to complete the post-sailing test on the same day as the pre-sailing test. They completed the post-sailing test one week later. The cancellation of the racing due to lack of wind on the day that the pre-sailing test was completed meant that completion of the afternoon test would have been pointless. This may have had the effect of reducing the number of correctly answered questions in the test due to the absence of any temporary learning effect from the morning test (as previously discussed).

Finally, the subjective measure of the effort expended by participants in completing the test indicated that effort level variations were not responsible for differences between the tests.

At the outset of this study it was argued that the failure of previous research to establish a consistent pattern of results in examining the after-effects of physical exertion may have been due to the nature of the cognitive tasks employed; specifically that the task durations were insufficient. Having successfully manipulated the physical tiredness, and controlled for the other intervening variables in the model, it was expected that performance on the 30-minute GTT would be worse after a day of sailing than after a day ashore. This was not supported. It appears that the performance decrementing factors of moderate physical tiredness either do not decrement cognitive performance on a test such as the GTT, are more subtle than the present study has been able to detect, or are not captured in the proposed model.

6.6 Implications for Racing Sailors

America's Cup yacht racing has been likened to a game of chess played with a rugby team, using Formula One racing cars. Certainly, yacht racing is a cognitively complex sport. A successful sailor must to attend to, process, successfully interpret, make decisions about, and respond to a large number of rapidly and subtly changing cues. It appears that the sailors involved the present study remain unaffected, in cognitive terms, by the physical demands of the sport.

The long- and short-term practice effects demonstrated in the study, together with the circadian effect indicate that both practice and "warming up" could significantly improve cognitive performance in the race. The improved performance on the second day of tests, in excess of a week after the first testing day, indicates that practising complex tasks improves the speed at which the answers can be correctly obtained. Further, every test showed an improvement in accuracy as the test progressed. These two results indicate that both long-term practice and warm-ups for each race can be cognitively beneficial. Additionally, the increase in the number of correctly answered questions in the afternoon, as compared to the morning tests indicates that effort needs to be made to overcome the circadian effect that seems to restrict performance in the mornings.

6.7 Problems and Future Directions

The use of a long-duration complex test to reduce the effects of motivation and interest resulted in some administration problems. The participants at times appeared to be bored, particularly when the group of young people involved would have preferred to be sailing. On one occasion (the morning of a lay-day) this proved to be disruptive. There was considerable disturbance from a small and restless group that required the test administrator to intervene and reduce the noise and distraction. The four sets of tests excluded from the analysis on the basis of low accuracy scores may be further evidence of the disruptive effects of boredom.

Due to an error in the administration of the post-lay-day test for a group of 12 participants, the time at which participants indicated the number of questions they had completed was at the five-minute point, rather than at the three-minute point. This was corrected by estimating the number of questions at three minutes by multiplying the 5-minute results by .6. Using the corrected information, the ratio of questions attempted in the first three minutes to the number of questions attempted in the total test was the same for both tests administered to that group on the day. Further, the rate of correctly answering questions did not differ amongst the phases. It is unlikely that estimating of the number of questions completed would have effected the overall results.

The GTT and the subjective measures employed in the study have little in the way of validation support. With respect to the GTT, there is a particular need for the test to be proven as a reliable measure of performance on a logical reasoning task. Although some limited work in this area was completed in the pilot study, further research support would be needed before there would be sufficient assurance of this. Notwithstanding there is some evidence that the GTT and the BLRT differ in terms of the constructs being measured, the GTT does not rely for its effectiveness on a clear understanding of these constructs. All that is required is assurance that the GTT is a reliable measure of cognitive performance on a complex cognitive task. It was the comparisons among the test occasions, and not the absolute values of the results, that were important for the study.

The same is not true of the subjective measures. These were developed for this study and are empirically unsupported, in terms of both reliability and validity. Furthermore, it is important that the tools do measure what they purport to measure; physical tiredness, mental tiredness, and effort expended. The measures used in the study were employed principally for their simplicity. The use of better supported tools in future research would increase the level of assurance that manipulations and controls have functioned as expected.

The measures used in the study, the accuracy of responding and the total number of correctly answered questions, were affected in different ways by the variables. The total number of questions answered appeared to be more sensitive to practice and circadian effects that occurred as a result of each administration. Accuracy, on the other hand, was influenced by practice effects that occurred during each administration, and was not influenced at all by the longer-term effects. There is a case for including both measures in future studies.

It appears that, although sleep loss studies appear to find a consistent pattern of cognitive impairment, the same cannot be said for the relationship between the after-effects of physical exertion and cognitive performance. It seems that it is the circumstances from which tiredness arose, rather than simply the feeling of tiredness, that will determine whether cognitive performance will be impaired.

References

Allison, D. B., Gorman, B. S., & Primavera, L. H. (1993). Some of the most common questions asked of statistical consultants: Our favourite responses and recommended readings. *Genetic, Social and General Psychology Monographs, 119,* 155-185.

Andersen, M. B., & Stoove, M. A. (1998). The sanctity of *p* < .05 obfuscates good stuff: A comment on Kerr and Goss. *Journal of Applied Sport Psychology, 10,* 168-173.

Anshel, M. H. (1997). *Sport psychology: From theory to practice* (3rd ed.). Scottsdale, AZ: Gorsuch Scarisbrick.

Baddeley, A. D. (1968). A 3 minute reasoning test based on grammatical transformation. *Psychonomic Science, 10,* 341.

Bartley, S. H. (1976). What do we call fatigue? In E. Simonson, & P. C. Weiser (Eds.), *Psychological aspects and physiological correlates of work and fatigue* (pp. 409-414). Springfield, IL: Charles C Thomas.

Blagrove, M., Alexander, C., & Horne, J. A. (1995). The effects of chronic sleep reduction on the performance of cognitive tasks sensitive to sleep deprivation. *Applied Cognitive Psychology, 9,* 21-40.

Bohnen, H. G. M., & Gaillard, A. W. K. (1994). The effects of sleep loss in a combined tracking and time estimation task. *Ergonomics, 37,* 1021-1030.

Danaher, J. (1995). Fatigue symposium proceedings: November 1-2 1995. *Opening remarks.* Symposium conducted at the meeting of the National Transport Safety Board and NASA AMES Research Center.

Davey, C. P. (1973). Physical exertion and mental performance. *Ergonomics, 16,* 595-599.

Dawson, S., & Petrie, K. J. (1996). Report on flightcrew fatigue on the FRA 0012 tour of duty. Unpublished manuscript.

Dinges, D. F. (1990). *The nature of subtle fatigue effects in long-haul crews*. Paper presented at the meeting of the Flight Safety Foundation, Rome.

Féry, Y. A., Ferry, A., Vom Hofe, A., & Rieu, M. (1997). Effect of physical exhaustion on cognitive functioning. *Perceptual and Motor Skills, 84,* 291-298.

Hancock, S., & McNaughton, L. (1986). Effects of fatigue on ability to process visual information by experienced orienteers. *Perceptual and Motor Skills, 62,* 491-498.

Hanton, S., & Jones, G. (1999). The effects of a multimodal intervention program on performers: II. Training the butterflies to fly in formation. *The Sport Psychologist, 13,* 22-41.

Hockey, R. (1983). *Stress and fatigue in human performance*. Chichester, England: Wiley.

Holding, R. (1983). Fatigue. In R. Hockey (Ed.), *Stress and fatigue in human performance* (pp 145-167). Chichester, England: Wiley.

Horne, J. A., & Pettit, A. N. (1985). High incentive effects on vigilance performance during 72 hours of total sleep deprivation. *Acta Psychologica, 58,* 123-139.

Howell, D. C. (1997). *Statistical methods for psychology* (4th ed.). Belmont, CA: Duxbury Press.

Jacques, C. H. M., Lynch, J. C., & Samkoff, J. S. (1990). The effects of sleep loss on cognitive performance of resident physicians. *Journal of Family Practice, 30,* 223-229.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Kremer, J. M. D., & Scully, D. M. (1994). *Psychology in sport.* London: Taylor & Francis.

Krueger, G. P. (1989). Sustained work, fatigue, sleep loss and performance: A review of the issues. *Work and Stress, 3,* 129-141.

Linde, L., & Bergstrom, M. (1992). The effect of one night without sleep on problem-solving and immediate recall. *Psychological Research, 54,* 127-136.

Lingenfelser, T. H., Kaschel, R., Weber, A., Zaiser-Kaschel, H., Jakober, B., & Küper, J. (1994). Young hospital doctors after night duty: Their task-specific cognitive status and emotional condition. *Medical Education, 28,* 566-572.

May, J., & Kline, P. (1987). Measuring the effects upon cognitive abilities of sleep loss during continuous operations. *British Journal of Psychology, 78,* 443-455.

McNaughten, D., & Gabbard C. (1993). Physical exertion and immediate mental performance of sixth-grade children. *Perceptual and Motor Skills, 77,* 1155-1159.

Microsoft Excel 97 [Computer software]. (1996). Redmond, WA: Microsoft Corporation.

Raviv, S., & Low, M. (1990). Influence of physical activity on concentration among junior high-school students. *Perceptual and Motor Skills, 70,* 67-74.

Rosekind, M. R., Gander, P. H., Miller, D. L., Gregory, K. B., McNally, K. L., Smith, R. M., & Lebacqz, J. V. (1992). Pilot fatigue, sleep and circadian rhythms: NASA fatigue counterprogram. *Aviation Safety Journal, 3,* 20-25.

Rust, J. O., & Barnard, D. (1979). WAIS verbal-performance differences among elderly when controlling for fatigue. *Psychological Reports, 44,* 489-490.

Ryman, D. H., Naitoh, P., & Englund, C. E. (1985). Decrements in logical reasoning performance under conditions of sleep loss and physical exercise: The factor of sentence complexity. *Perceptual and Motor Skills, 61,* 1179-1188.

Shiffren, R. M., & Schnieder, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84,* 127-190.

Soetens, E., Hueting, J., & Wauters, F. (1992). Traces of fatigue in an attention task. *Bulletin of the Psychonomic Society, 30,* 97-100.

Sothmann, M. S., Hart, B. A., & Horn, T. S. (1992). Sympathetic nervous system and behavioral responses to stress following exercise training. *Physiology & Behavior, 51,* 1097-1103.

Sparrow, W. A., & Wright, B. J. (1993). Effect of physical exercise on the performance of cognitive tasks. *Perceptual and Motor Skills, 77,* 675-679.

Tabachnick, B. G., & Fidel, L. S. (1989). *Using multivariate statistics* (2nd ed.). New York: Harper Collins.

Tactics and strategy simulator [Computer software]. (1992). Haddam, CT: Posey Yacht Design.

The SAS System (Release 6.12) [Computer software]. (1989). Cary, NC: SAS Institute Inc.

Tomporowski, P. D., & Ellis, N. R. (1986). Effects of exercise on cognitive processes: A review. *Psychological Bulletin, 99,* 338-346.

Webster's universal dictionary and thesaurus. (1993). Montreal: Tormont.

Wendt, H. W., & Palmerton, P. R. (1976). Motivation, values, and chronobehavioral aspects of fatigue. In E. Simonson, & P. C. Weiser (Eds.), *Psychological aspects and physiological correlates of work and fatigue* (pp. 285-335). Springfield, IL: Charles C Thomas.

Zervas, Y. (1990). Effect of a physical exercise session on verbal, visuospatial, and numerical ability. *Perceptual and Motor Skills, 71,* 379-383.

Zervas, Y., Danis, A., & Klissouras, V. (1991). Influence of physical exertion on mental performance with reference to training. *Perceptual and Motor Skills, 72,* 1512-1221.

Appendix A

Test #	

Grammatical Transformation Test

(adapted from Baddeley (1968) and Blagrove, Alexander and Horne (1995))

Instructions

In the following test there are a number of short sentences each followed by a pair of letters. The sentences refer to the letters. For some sentences you must decide if the sentence correctly describes the two letters, and put a tick under 'True' or 'False' accordingly. Some other sentences are conditional instructions: if the conditional sentence correctly describes the two letters then place a tick as directed by the sentence. If it does not describe the two letters correctly do not put a tick at all. The following are examples:

	True	False
M is followed by C - MC	√	
C precedes M - MC		√
If M is followed by C then tick False - CM		
If M is smaller than C then tick False - Cm		√

You now have 30 minutes to work through as much of the test as you can. Your final score will be the number right minus the number wrong.

Please wait for the administrator to tell you to start.

OFFICE USE ONLY

S/B	*Ver #*	*Pa#*	*Con #*
		#	√
Total			
0 - 3			
0 - 10			
20 - 30			
Active			
Passive			

		True	***False***
1	If M is not followed by C then tick False - CM		
2	If M is preceded by C then tick True - MC		
3	A is preceded by B - AB		
4	If M precedes C then tick True - CM		
5	B precedes A - AB		
6	B is not preceded by A - BA		
7	If C is larger than M then tick False - Cm		
8	B precedes A - AB		
9	If M does not precede C then tick False - MC		
10	B is preceded by A - AB		
11	If C follows M then tick True - CM		
12	If M does not follow C then tick True - MC		
13	If C does not follow M then tick True - MC		
14	B does not follow A - BA		
15	A does not follow B - BA		
16	If C is preceded by M then tick False - CM		
17	If M is not to the right of C then tick False - MC		
18	If C precedes M then tick False - MC		
19	If C is not preceded by M then tick False - MC		
20	B is not preceded by A - AB		
21	If C precedes M then tick True - MC		
22	If M is to the left of C then tick False - CM		
23	A does not precede B - BA		
24	A is not preceded by B - BA		
25	If M is to the left of C then tick False - MC		
26	A precedes B - BA		
27	A does not precede B - BA		
28	If M is preceded by C then tick True - CM		
29	If M is followed by C then tick False - CM		
30	If C is not preceded by M then tick True - MC		
31	If C is not smaller than M then tick False - Cm		
32	B does not precede A - AB		
33	If C does not precede M then tick False - CM		
34	If C is not to the left of M then tick False - MC		
35	A follows B - AB		
36	If M is followed by C then tick True - CM		
37	B is followed by A - AB		
38	If M follows C then tick True - MC		
39	If C is not to the left of M then tick True - CM		
40	A is preceded by B - BA		
41	If C is preceded by M then tick True - MC		
42	If C is not followed by M then tick False - CM		
43	A is followed by B - BA		
44	If M follows C then tick False - CM		
45	B is preceded by A - BA		
46	If M is to the right of C then tick True - CM		
47	B is followed by A - BA		
48	If M is larger than C then tick True - Mc		
49	If M is larger than C then tick False - Cm		
50	If M is not to the left of C then tick False - MC		

Appendix B

Grammatical Transformation Test - Verbal Instructions

(adapted from Baddeley (1968) and Blagrove, Alexander and Horne (1995))

Before commencement:

Introduction:

- Please make sure you have the correct test from your test envelope. This is test number *[appropriate test number]*. The number is written on the top of your instruction sheet.
- Please do not start yet.

Subjective Fatigue:

- First, I would like you to indicate how physically and mentally tired you feel at the moment.
- On a scale of 1 to 7, select the number that best indicates your physical tiredness - with:
  - 1 meaning your body feels fully refreshed and ready for anything
  - 7 meaning you feel completely exhausted and unable to attempt any physical activity .. and ..
  - 4 being in the middle between those two extremes.
- Please write this number under the "Test number" box at the top right corner of the front page.
- Now you need to do the same thing for mental tiredness:
  - 1 meaning your mind is totally clear and fresh and ready for demanding thought
  - 7 meaning you feel completely mentally exhausted with your mind not able to think clearly at all .. and ..
  - 4 being in the middle
- Please write this number (between 1 and 7) under the *physical* tiredness number you have just written.

Pre-test Instructions:

- It is important you give the test your best effort. Please treat it seriously.

- You will have 30 minutes to work on the test to complete as many questions as you can. Please concentrate on being as accurate as possible.
- At various times during the test you will be asked to mark in the margin the question you are currently working on. I will tell you when to do this.
- You now have a few minutes to read through the instructions (again) and ask any questions.
- Once you have completed reading the instructions, please wait for me to tell you to start.

---Pause for reading time until it appears everyone is ready.

- You all understand that "precedes" means "comes before"?
- If you think you will run out of questions before the end of the time, please raise your hand to let me know when you get near the end, and I will give you another sheet of questions.
- Are there any (other) questions? *[Answer those questions related to clarification of the instructions only. All other questions should be reserved until after the completion of the test.]*
- Okay you can start. You have 30 minutes.

At exactly 3 minutes into the test:

- Can you please pause and write a "1" in the left hand margin next to the number of the question you are currently working on then continue the test.

At exactly 10 minutes into the test:

- Can you please pause and write a "2" in the left hand margin next to the number of the question you are currently working on then continue the test.

At exactly 20 minutes into the test:

- Can you please pause and write a "3" in the left hand margin next to the number of the question you are currently working on then continue the test.
- You have 10 minutes left.

At the end of the time (30 minutes):

- Could you please stop writing.

- Can you please write a "4" in the left hand margin next to the number of the last question you worked on, whether or not you finished it.

Subjective Effort:

- Before I collect in the tests, I would like you to assess how hard you tried with this test compared to the practice test you did a week *(or more)* ago. Remember this is anonymous so please give an honest assessment.
- On a scale of 1 to 7, select the number that best indicates your how hard you tried in this test compared to the original practice test - with:
    - 1 meaning you tried way harder than on the practice test
    - 7 meaning you barely tried at all with this test .. and ..
    - 4 meaning you tried just as hard, but no harder, than you did with the practice test.
- Please write this number (between 1 and 7) at the very bottom right hand corner of the front page.

Wrap Up:

- When you have done this I need to make sure I get all the tests in. So:
    - please bring your finished test to me
    - please close up your envelope with the unused tests in it, then put your envelope and your pen back on the *(collection)* table.
- Thank you for your effort.