Test Length and Cognitive Fatigue:
Phase II: Empirical Examination of Performance Effects and Examinee Reactions

Final Report

to

College Board
45 Columbus Avenue
New York, NY 10023

Phillip L. Ackerman, Ph.D.
Ruth Kanfer, Ph.D.

School of Psychology
Georgia Institute of Technology

November 7, 2007

Executive Summary

Background. In the first phase of this project, we conducted an extensive literature review of the domain of cognitive fatigue, especially in the context of the kinds of activities associated with ability assessment. The review indicated that a wide array of effects are found in similar situations, ranging from no appreciable changes to mean performance, to decreases or even significant increases in performance over the course of extended cognitively-demanding tasks. In contrast to the wide array of performance effects, there is a general pattern of subjective reports of increasing cognitive fatigue as time-on-task increases. As a result, the typical result of such investigations is to find a dissociation between reports of subjective fatigue and objective performance stability or change. In addition, some empirical studies found suggestive data on patterns of individual differences in reactions to cognitively-demanding tasks over time.

We concluded the review with the following:

"Because high-stakes cognitive testing situations like the SAT represent a unique combination of task characteristics and individual characteristics (both trait and state), it is impossible to determine the performance and subjective fatigue effects without an empirical investigation that focuses specifically on these issues. It is recommended that a study of both performance effects and subjective reactions be conducted. The study needs to take account of individual differences in affective and conative trait factors, along with individual differences in state variables and prior experiences, in order to determine which individuals are likely to increase performance, decrease performance, or be unaffected by testing of shorter or longer durations. In addition, assessment of subjective fatigue before and after tests of different durations should be assessed as a criterion variable for subjective reactions to the testing situation."

Phase II. In this phase of the project, we conducted an empirical examination of the performance and subjective report aspects of SAT testing across different conditions of total-time-on-test.

1. Two hundred and thirty-nine first year university students were recruited.
2. Participants were first given an assessment of traits (e.g., personality, interest, motivation, self-concept) that might account for variance in either SAT performance and/or subjective perceptions of fatigue. The assessments were done several days prior to the first SAT testing session.
3. In a within-participant design, participants completed the SAT under three different conditions: A Standard test time condition (total time 4½ hr), a Short test time condition (total time 3½ hr) and a Long test time condition (total time 5½ hr). Condition order and test order were fully counterbalanced.
4. Participants completed pretest, interim, and posttest assessments of subjective fatigue and related scales of negative affect, positive affect, positive motivation, confidence

and self-efficacy.

Results

1. Overall SAT equivalence scores from each test condition were compared to archival College Board SAT records for the participants, and across the test conditions. Average SAT equivalence scores (based on Verbal + Quantitative composites) were 20 points lower than performance in the high-stakes situation. However, 48% of the participants had scores that equaled or exceeded their SAT archive scores.
2. Performance in the longer test conditions, on average, exceeded that of the shorter test conditions. Mean performance in the 4½ hr condition was 13 points higher than in the 3½ hr condition. Mean performance in the 5½ hr condition was 15 points higher than in the 4½ hr condition.
4. There were significant and substantial increases in subjective fatigue with time-on-test in *each* condition.
5. Posttest subjective fatigue in the Short and Standard conditions correlated $r = .70, .68$ with pretest subjective fatigue, respectively. In the Long condition, posttest subjective fatigue correlated $r = .48$ with pretest subjective fatigue. Posttest subjective fatigue in the Short condition correlated $r = .72$ with posttest subjective fatigue in the Standard condition and $r = .62$ with posttest subjective fatigue in the Long condition.
6. In the Short and Standard conditions, subjective fatigue correlated less than $r = .20$ with SAT performance. In the Long condition, subjective fatigue correlated no more than $r = .25$ with SAT performance.
7. In the prediction of SAT performance, subjective fatigue did not account for incremental variance beyond trait complexes and self-concept/self-estimates of abilities in either the Short or Standard conditions. In the Long condition, subjective fatigue accounted for from 4-6% of the variance in final-hour SAT performance, above that accounted for by trait complexes and self-concept/self-estimates of abilities.
8. Roughly one-third of the variance in subjective fatigue could be accounted for by individual differences in stable trait complex measures, collected several days *prior* to testing. Also, a similarly large portion of the posttest subjective fatigue variance was accounted for by pretest subjective fatigue (assessed prior to the test on the test day). In the Standard (4½ hr) SAT session, 52 percent of the variance in posttest subjective fatigue was accounted for by measures administered *prior* to the test.

Acknowledgements

The authors wish to acknowledge the support and able assistance of the College Board in sponsoring this research project, and in providing test forms and scoring. In particular, we are deeply grateful for the support provided by Wayne Camara, Krista Mattern, and Glenn Milewski. In addition, we wish to acknowledge the assistance of our lab staff in recruitment, administering the tests, coding data and other critical activities. In particular, we acknowledge the efforts of Stacey Wolman, Sunni Haag, and the rest of the Knowledge & Skill Lab staff at Georgia Tech.

Background and Introduction

A longstanding issue for ability/aptitude/achievement testing concerns examinee fatigue, especially when tests are longer than an hour or so. The issue of cognitive fatigue during testing has been raised at various times during the last hundred years of modern assessment research and practice. In fact, the earliest examples of group ability assessment (Ebbinghaus, 1896-1897) were not actually aimed at a direct assessment of individual differences in ability, but rather were developed to assess the effects of cognitive fatigue (in terms of the length of the school day) on attention and performance. An early application of ability testing for university selection, for example, was characterized as an assessment of both ability *and* perseverance.

There is a substantial older research literature on the performance effects associated with increasing the amount of time required by continuous demands on cognitive processing. However, until the first phase of our project, there had been no attempt to review or integrate research conducted since about 40 years ago. There has been no clear picture of the performance effects associated with test times that run as long as four hours or more. In addition, performance effects are only one part of the overall consideration of test length issues. The other significant aspect of extended testing has to do with examinee reactions. Although there have been a few studies investigating examinee reactions to test content or the effects of test-taking motivation in general, there has been minimal empirical research that would directly relate to examinee reactions to standardized testing of different lengths. In addition, there is little or no empirical research that has specifically focused on the identification of which traits may be associated with adverse examinee reactions to extended test lengths.

This report reflects the second phase of a two-phase approach to understanding the performance and examinee reaction effects associated with extended testing times for cognitive assessments. The first part of the proposed project involved an integrated review of the effects of extended time-on-task in cognitively demanding tasks on performance and examinee reactions (see Ackerman & Kanfer, 2006). This (second) phase of the project involves an empirical investigation of the determinants (e.g., ability, personality, motivation, and other determinants) of performance effects and examinee reactions to tests of different lengths. The goal of Phase II of this project was to: (a) provide an assessment of mean fatigue effects (performance and subjective) in testing of three different total times-on-task, and (b) identify patterns of individual performance and subjective fatigue effects in the context of the different testing times. It was expected that the outcome of the project will be to provide reliable and valid assessment of the effects of extended testing time on cognitive fatigue, and provide information that might help potential examinees better prepare for the exam in a way that may attenuate or eliminate these effects.

Phase I. Review of the Literature

In the first phase of this project, we reviewed an extensive body of literature on cognitive fatigue, and a variety of related research domains, including task characteristics, individual characteristics [e.g., personality and motivation], and situational characteristics [e.g., the physical testing environment, mood, attitudes toward testing, etc.]). Our report fully documents this review, but to summarize from the report:

"The literature reviewed to date suggests that mean decrements in cognitive test performance are likely to be relatively small over the course of extended time-on-task, up to several hours of cognitive ability testing. In fact, there are extant data that demonstrate either no mean effects, small increases, or small decreases in mean performance under extended time completing cognitively involving tasks. Mean increases in performance have often been found when incomplete task practice results in continued improvement with additional time-on-task. Mean decreases have been found less often, but they are most often associated with several key task situation characteristics, namely: (a) high demands on intellectual functioning, (b) high demands on attention to detail, (c) low tolerance for errors; (d) lack of knowledge of results/feedback, (e) high levels of time pressure (speeded tests); (f) lack of breaks, (g) high performance costs of distractions, (h) presence of environmental stressors (e.g., noise, poor illumination, heat), (i) high-stakes testing, (j) tasks that are not intrinsically interesting or enjoyable, (k) high levels of visual demands for reading (e.g., small text); (l) low probability of success; and (m) particular kinds of test content (verbal content showing more rapid fatigue effects than math content). Although the effects of any one of these task characteristics may result in no significant mean decrements in performance over extended testing time, a combination of several of these characteristics might be expected to have a more substantial overall effect on performance.

The literature on individual differences in susceptibility to fatigue effects is more suggestive, but there have been several salient findings. The extant research suggests that when individuals become fatigued (which is only a characteristic of some portion of the population, depending on the task situation), some individuals respond by increasing task effort (overactivity), and therefore maintain or increase task performance; some individuals respond by decreasing task effort (withdrawal), and thus show a decrease in task performance; and still other individuals first respond by increasing effort, followed by a later withdrawal of effort. Both trait (i.e., ability, personality, motivation), state (sleep deprivation, recent meals, drugs), and experiential (e.g., prior exposure to similar situations) variables appear to influence the subjective judgements and strategies adopted by individuals under cognitively fatiguing conditions.

Subjective fatigue follows a pattern that is only partly linked to objective performance effects. In general, mean subjective judgements of fatigue tend to precede performance effects, that is, individuals typically *feel* fatigued before they actually show a performance decrement. Subjective fatigue under cognitive ability testing has a substantial undifferentiated component (meaning that individuals report many different aspects of fatigue, including drowsiness and dullness, difficulty in concentration, and physical fatigue of body parts). However, there is a reasonably good hierarchical theoretical framework for parsing the underlying structure of subjective fatigue, one that allows for the determination of both general and specific effects of the task situation on perceived fatigue." (Ackerman & Kanfer, 2006).

Phase II. Empirical examination of test length and cognitive fatigue effects

The following task characteristics were identified in the literature as related to a greater or lesser extent in cognitive fatigue (in addition to the most critical issue of total continuous time-on-task), including:

(a) high demands on intellectual functioning
(b) high demands on attention to detail
(c) low tolerance for errors
(d) lack of knowledge of results/feedback,
(e) high levels of time pressure (speeded tests)
(f) high performance costs of distractions
(g) high-stakes testing
(h) tasks that are not intrinsically interesting or enjoyable
(i) low probability of success *[for some examinees]*
(j) particular kinds of test content (verbal content showing more rapid fatigue effects than math content)
(k) high levels of visual demands for reading (e.g., small text)
(l) presence of environmental stressors (e.g., noise, poor illumination, heat)
(m) lack of breaks

Of the 13 major task characteristics that have been shown to be involved to a greater or lesser extent in cognitive fatigue, the SAT has 9 or 10 such characteristics (the last three items do not generally apply to the SAT test situation). Adding the most critical issue of total time-on-task at 3 hr, 45 min (over a nearly 5 hour session), it appears that there is a substantial likelihood of some cognitive fatigue effects on performance, *at least for some individuals.* That is, extant research has indicated that mean effects of cognitive fatigue on performance (at least within a 3-6 hour total time-on-task) are often relatively small. However, mean performance levels appear to hide four types of examinee reactions to these cognitive demands, namely: overactivity, underactivity, no reaction, and overactivity followed by underactivity. Thus, depending on the proportion of examinees who show one of these patterns of responses over the course of testing, mean effects may not be seen (or may be reflected in small increases or decreases), but there may be, in fact, performance effects associated with cognitive fatigue at the individual or group level.

Several individual and situational characteristics have been shown to be related to performance effects under these kinds of task conditions, and it was our goal to evaluate these characteristics in the context of a prediction model for cognitive fatigue performance effects.

In addition to performance measures, the assessment of subjective fatigue is also an important consideration. Some subjective fatigue factors appear to be correlated with performance effects, but they are also important in terms of examinee reactions to the testing situation.

Study Overview

The general outline for the empirical study is as follows:

(1) Pretest assessment of trait predictors of performance and examinee reactions, by use of an extensive questionnaire of personality and other measures completed by the participant at home, up to two weeks prior to the first testing session.

(2) Test assessments of differing lengths. Tests of three different test lengths were administered, in a counterbalanced, *within-subjects* design. The test lengths were: Standard (standard SAT test length), Short (one hour shorter than Standard), and Long (one hour longer than Standard).

(3) Examinee reactions. Short assessments of examinee attitudes and reactions were made prior to, during, and immediately following the testing sessions.

Distal Predictors (Completed 1 to 2 weeks prior to first test session)
1. Personality/affect
2. Conation/motivation
3. Academic and Ability Self-concept, self-estimates of abilities
4. Caffeine intake
5. Interests

Proximal Predictors
1. Prior night sleep amount/quality
2. Pretest subjective fatigue, negative and positive affect, positive motivation, confidence, and self-efficacy.

Criteria & Other Correlates
1. SAT test performance (overall and during last 50 minutes of test session)
2. Subjective fatigue
3. Negative affect
4. Positive affect
5. Positive motivation
6. Subjective confidence
7. Self efficacy
8. Reports of effort change and strategy during testing session
9. Cognitive ability (SAT scores under high-stakes conditions)

Hypotheses

Based on the extant research from a variety of different fields, we expected that the level of *mean* performance on the cognitive measures will show relatively little difference between testing times of 3½ hours, 4½ hours, or 5½ hours. There were mixed data in the literature, however, on the specific nature of effects of test length on errors (and number of items attempted), which may be more reactive to fatigue effects than raw number correct performance. There are two basic elements to the expected results, with respect to individual differences -- namely: (a) distal and proximal predictors of cognitive fatigue, and (b) strategic reactions to cognitive fatigue. We expected that several trait-level variables (e.g., self-concept/self-estimates of ability, neuroticism/anxiety, achievement orientation, interest, etc.) will be related to performance effects (both positive and negative) and/or subjective reactions (though more strongly to subjective reactions), and that there will be some involvement of proximal variables (such as mood, attitudes toward the testing situation, prior night's sleep, self-efficacy, etc.). We also expected that there it will be possible to identify the four categories of individual strategies described above, and then to relate these strategy classifications to the distal and proximal measures. From these two elements, the expectation was that it will be possible to identify the most important individual characteristics that are related to subjective judgments of fatigue and performance effects.

Method

Participants

Freshman students were identified from institutional records. Letters of invitation were sent by regular mail and/or e-mail to students at the Georgia Institute of Technology, Georgia State University, and Emory University. A total of 251 students participated in the study. Data from twelve participants were eliminated from analysis. Six of the participants either failed to follow instructions or fell asleep repeatedly during testing. The remaining six students exerted minimal effort during testing (with average scores approximately 200-400 points below their College Board record scores across all three sessions of the study). For the remaining 239 participants, there were nearly equal numbers of men and women (124 men and 115 women), with an average age of 18.3 (sd = .53) and a range of ages from 17 to 20. The College Board record SAT scores (Math + Verbal) for the sample had a mean of 1249, sd = 165, and a range from 800 to 1600; indicating that this sample was overall above average in performance on the SAT compared to national norms, but showed no indication of ceiling or floor effects (only one participant had a score of 1600). Of the 239 participants, 38 completed only one testing session, 17 completed two of the three testing sessions, and 184 completed all three sessions. There were no significant differences in recorded SAT scores between those individuals who completed one, two or three sessions (MSE = 4302; $F(2,226)$ = .16, ns).

At-Home Questionnaire (AHQ)

The AHQ packet included several self-report scales designed to assess cognitive, affective, and conative traits that were expected to be related to test performance and self-ratings of attitudes and emotions during testing. The scales were designed to provide robust markers for a small (six or fewer) trait complex factors (e.g., see Ackerman, Bowen, Beier, & Kanfer, 2001). A list of all of the scales from the AHQ, along with the number of items, mean, *sd* and internal consistency reliability (α) are provided in **Table 1.** The measures were as follows:

International Personality Item Pool. Scales included: Extroversion, Need for Achievement (nAch), Cautiousness, Impulse Control, and Self-Discipline (Goldberg, 2005). Each scale was composed of 7 - 15 items that were balanced in terms of positive or negative statements.

Typical Intellectual Engagement (TIE). The 59-item TIE (Goff & Ackerman, 1992) was also administered. This scale was designed to assess an individual's pattern of interests, attitudes and behaviors regarding intellectually-demanding activities (α = .92, Ackerman et al., 2001; convergent validity estimates of .49 with verbal ability, Rolfhus & Ackerman, 1996).

Numerical Preferences. An adaptation of the Viswanathan's (1993) Preference for Numerical Information scale was administered. This scale is composed of 11 items and measures an individual's attraction or aversion to working with numerical or mathematical tasks.

Motivational Trait Questionnaire. The short-form of the Motivational Trait Questionnaire (MTQ; Kanfer & Ackerman, 2000; see also Heggestad & Kanfer, 2000; Kanfer & Heggestad, 1997) is a 48-item measure that contains six scales. The scales in this measure were composed of 6 to 10 items, and included assessments of Approach-oriented motivation (Desire to Learn, Mastery), Competitive Excellence (Other-referenced goals, Competitiveness), and Aversion-related motivational traits (Worry, Emotionality).

Motivated Strategies for Learning Questionnaire (MSLQ). The MSLQ is an 81 item self-report questionnaire. The MSLQ was designed to measure "motivation and use of learning strategies by college students" (Pintrich, Smith, Garcia, & McKeachie, 1993, p. 801). It has two sets of scales: Motivational Scales (Intrinsic Goal Orientation, Extrinsic Goal Orientation, Task Value, Control of Learning Beliefs, Self-Efficacy for Learning & Performance, Test Anxiety); and Learning Strategy Scales (Rehearsal, Elaboration, Organization, Critical Thinking, Metacognitive Self-Regulation, Time and Study Environmental Management, Effort Regulation, Peer Learning, and Help-Seeking).

State-Trait Anxiety Inventory. The 20-item Spielberger (1983) State-Trait Anxiety Inventory (STAI) was administered under "trait" instructions. It is designed to assess broad aspects of anxiety (in contrast to the school or testing settings).

Interests. Four scales (60 items) from The Unisex American College Testing Interest Inventory (UNIACT; Lamb & Prediger, 1981) were administered. The four interest themes assessed were identified by Holland (1973) as: Realistic, Investigative, Artistic, and Conventional.

NEO-FFI (NEO-Five Factor Inventory). The NEO-FFI (Costa & McCrae, 1992) was administered. The scales were composed of 12 items each and included: (1) Neuroticism, (2) Extroversion, (3) Openness to Experience, (4) Agreeableness, and (5) Conscientiousness. Responses were made on a 6-point Likert-type scale (1 = Very untrue of me to 6 = Very true of me).

Self-Consciousness Focus. Five items from the private self-consciousness scale (Fenigstein, Scheier, & Buss, 1975) were administered. These items refer to the way in which the participant monitors inner feelings and changes in mood.

Behavioral Inhibition System/Behavioral Activation System (BIS/BAS). This measure (Carver & White, 1994) has four scales, based on Gray's (1990) demarcation of the approach (activation) and avoidance (inhibition) systems. Four scales make up the BIS/BAS measure (BIS, BAS Reward Responsiveness, BAS Drive, and BAS Fun Seeking). The scales have from 4 to 7 items each.

Caffeine Consumption. A locally developed measure of typical caffeine consumption was administered. Questions on the scale pertained to how dependent the individual perceives himself/herself to be on caffeine (e.g., "If I have to wake up much earlier than usual, I will try to make sure I have a beverage with caffeine." and "If I don't have a caffeinated beverage early in the day, I tend to be irritable.")

Morningness/Eveningness. An adapted version of the Horne-Östberg (1976) Morningness-Eveningness Questionnaire was administered. This 19-item scale includes questions about an individual's preferences for sleeping and waking times, and time-of-day preferences for various activities.

Self-Concept. Two scales of academic self concept were administered (e.g., see Ackerman & Beier, 2006; Ackerman & Wolman, 2007). The scales (Verbal and Math) were composed of 6 items each.

Self-Estimates of Ability. Self-estimates of Verbal ability, Math ability, Spatial ability, Memory ability, and a single question of self-estimated general intelligence were administered. These items were administered with a format that required participants to indicate their relative standing, in percentile terms.

Additional Questions. In the final section of the AHQ, several individual questions were included about prior experience (e.g., “How many times did you take the SAT?” “How many practice SAT tests did you complete”) and attitudes (“I thought that the math/verbal/essay section of the SAT was difficult”).

SAT Tests

Three versions of the SAT battery of tests were provided by the College Board, designated: “Short,” “Standard” and “Long”. The tests were based on disclosed items from previous SAT national administrations. Sections were counterbalanced, so that there were three different forms of each test, for a total of nine different test forms. The sections of the tests are summarized in the context of the overall procedure, in **Table 2**. As indicated in the table, the first 2½ hours of each test session were equivalent, in terms of the questionnaires and test parts. Each test battery started with an Essay section, and then was followed by alternating subtests of verbal, math, or writing sections. Equivalent 50-minute segments (one 20 min Verbal section, one 20 min Math section, and a 10 min Writing section) were administered in the final 50 min of each condition. In addition, an equivalent segment was administered in the penultimate 50 min of the Long condition.

Scoring of the SAT tests was completed by the College Board subsequent to data collection, using their formulas and scaling procedures. For each condition, SAT equivalent scores were computed from overall performance. In addition for the last 50-min segment in each condition (and the penultimate 50 min segment in the Long condition), component raw scores were computed, based on the College Board’s scoring algorithms.

Subjective Reports During Testing

Subjective reports were obtained at the start of each test session (before any SAT test items were administered). Interim questionnaires (with identical items to the pretest questionnaire) were administered after 2½ hour of each session. Additional retests of the interim questionnaires were administered after 3½ hour of testing in the Standard and Long conditions, and after 4½ hour of testing in the Long condition only. Posttest questionnaires (identical to the pretest and interim questionnaires, except for rephrasing of confidence-type questions to be retrospective) were administered immediately after the last SAT section, which came at 3½ hours in the Short condition, 4½ hours in the Standard condition, and 5½ hours in the Long condition. The posttest questionnaire in the Short condition was administered at the same time-in-session as the second interim questionnaires in the Standard and Long conditions, and the posttest questionnaire in the Standard condition was administered at the same time-in-session as the third interim questionnaire in the Long condition.

Pretest, Interim, and Posttest Questionnaires

A 52-item measure of state affect, emotions, and self-efficacy (from the Positive and Negative Affect Schedule, Watson, Clark, & Tellegen, 1988; Profile of Mood States, McNair, Lorr, & Droppleman, 2003; and locally developed items) was administered before the start of SAT testing in each condition (“Pretest”), periodically during the course of testing (“Interim” questionnaires at 2½ hr, 3½ hr, and 4½ hour in the Long condition, at 2½ and 3½ hr in the

Standard condition, and only at 2½ hr in the Short condition), and after the final SAT section in each condition. The questionnaire included scales of Subjective Fatigue (12 items, e.g., "worn out" "exhausted"), Negative Affect (12 items; e.g., "uneasy" "tense"), Positive Affect (7 items; e.g., "enthusiastic" "excited"), Positive Motivation (5 items; e.g., "I am ["was" on the posttest] motivated to get a high score on this test" "I will try ["tried" on the posttest] to work to my maximum potential"), and Confidence (7 items; "I am confident I will perform well ["did perform well," on the posttest] on this test" "I will have ["had" on the posttest] difficulty concentrating on this test" [Reverse-scored]). A list of the scales from these questionnaires, along with the number of items, mean, *sd* and internal consistency reliability (α) are provided in **Table 3.**

Additional questions at end of each session.

At the end of the posttest for each session, an additional set of questions were included. These questions pertained to the individual's report of increasing or decreasing effort during the course of the session, and whether they felt that the test was more or less fatiguing than taking a course final exam.

Subjective reports after final testing session.

At the end of the posttest on the third session (after completion of all SAT testing), a final set of questions were administered. These questions related to the participants' attitudes toward the SAT and their experiences with the SAT.

Archival Data

Subsequent to the completion of the study, we obtained archival records of the participants' high school cumulative grade point average (GPA) and their first-year university GPA. In addition, archival SAT scores, from the tests taken during high school (i.e., SAT tests administered under a high-stakes testing situation) were obtained from College Board records.

Procedure

Testing took place on three consecutive Saturday mornings. In all conditions, participants were instructed to arrive a few minutes before the official 8 a.m. start of testing. (On the first day of testing for each group, consent and academic release forms were checked, IDs were checked against the master list, and at-home questionnaires were turned in.) At 8:00 a.m. the doors to the classroom were closed and the procedure started. All instructions, test times and breaks were controlled by pre-recorded CDs presented over a public address system.

At the beginning of the study (in the consent form), participants were told that they would receive a bonus of $25 each if their SAT equivalence scores met or exceeded the archival SAT test scores (taken under high-stakes conditions). In addition, the participants were told that the five participants with the largest increase in scores in comparison to the archival scores would receive an additional $100 bonus.

The layout for each of the three conditions is shown in **Table 2**. First, the pretest questionnaire was completed, followed by general instructions to the participants. The first test was the Essay Test. The essay test was followed by alternating tests of verbal and math content. The first 5-min break came after one hour elapsed time in the study. The break was followed by an additional 75 min of testing, followed by a second 5-min break. After the break, the participants completed the first Interim Questionnaire (which was identical to the pretest questionnaire). At this point, 2½ hr of time had elapsed.

For the Short condition, the questionnaire was followed by a module that consisted of: a Verbal test (20 min), a Math test (20 min) and a Writing test (10 min). The test module was followed by the posttest questionnaire (which had overlapping questions with the Pretest and Interim Questionnaires, but also had some additional posttest questions). For the Standard and Long conditions, the questionnaire was followed by a Writing test (25 min) and a Math test (25 min), along with a third 5-min break and another Interim Questionnaire, at which point approximately 3½ hr of time in the classroom had elapsed. Next, the Standard condition followed the same final 50 min module of testing that was administered in the final portion of the Short condition. The Long condition followed the same procedure, except instead of a Posttest Questionnaire at the end of the module, participants were given a fourth 5-min break and another Interim Questionnaire. Finally, the Long condition participants completed an additional 50 min module of testing identical to the last 50 min of the Standard and Short conditions, followed by the Posttest questionnaire. Total elapsed time for the study was 3½ hr for the Short testing condition, 4½ hr for the Standard testing condition, and 5½ hr for the Long testing condition. At the end of the third session for each participant, the participants completed a post-experiment questionnaire (a final page to the last posttest questionnaire), then they were debriefed and compensated $150 for their participation.

Results

Performance Scores.

Two sets of scores represent the critical criterion data for the experiment. The first set of scores is a combined (verbal + math) SAT “Equivalence” score for each of the three sessions. The SAT Equivalence scores were derived by College Board staff, based on the individual test part scoring algorithms and formulas derived from the original test administrations for each of the three sessions. The second set of scores are from the last 50 min of each testing session (and the last two 50 min sections of the Long condition testing session). For each of these, an equivalent set of three tests were administered (one 20 min Verbal test, one 20 min Math test, and one 10 min multiple-choice Writing test). Each of these scores was the total number of items answered correctly minus a fraction (.25) for the total number of items answered incorrectly.

Test of Counterbalancing/Order Effects

A comparison was made of average SAT equivalence scores (Verbal + Math) for each of the three conditions (Short, Standard, and Long), based on whether the participants first completed the Short, Standard, or Long sessions. For the three comparisons ($F(2, 200) = .64$, $MSE = 19436$, $F(2,206) = .81$, $MSE = 24872$, and $F(2, 208) = 1.10$, $MSE = 34551$, for Short, Standard, and Long sessions, respectively), none indicated a significant effect of session order. Therefore, the data for different orders were combined for the subsequent analyses.

SAT Equivalence Scores

Comparison with College Board Records. The first analysis for the SAT equivalence scores was a comparison with CB records, to evaluate the overall difference between the high-stakes testing situation (CB records) and the low-stakes testing of the laboratory experiment. SAT Equivalence performance across the three sessions yielded a mean of 1225 (verbal + quantitative), and $sd = 170.30$. From the participants with matching CB records, SAT performance [verbal + quantitative] under high-stakes testing was Mean = 1245, $sd = 170.62$. A paired *t*-test indicated that the mean performance under high stakes testing was 19.34 points higher than under the laboratory testing conditions, which was significant ($t\ (175) = -3.41$, $p < .01$). However, 48% of the participants obtained average SAT equivalence scores that equalled or exceeded their CB record scores. The range of differences in comparison to the archival SAT scores, was from a drop of -253 points to an increase of 146 points. (Of the 52% that had lower scores in the study, two-thirds of the participants had scores that were less than or equal to 100 point difference [verbal + quantitative].) For those participants who did not obtain scores equal to their SAT archival test scores, the reasons for the differences were logically due to some combination of the following: (a) those participants were not sufficiently motivated to perform well on the tests; (b) that the conditions of testing had a negative effect on their performance; or (c) that the decline in performance was the result of regression to the mean.

SAT Equivalence Scores by Session. A comparison was made across the three sessions for overall SAT equivalence scores. For the Short session, SAT equivalence scores were $M = 1209$, $sd = 173$; for the Standard session, $M = 1222$, $sd = 174$; and for the Long session $M = 1237$, $sd = 177$. An ANOVA performed on these scores indicated a significant effect of session ($F\ (2, 364) = 6.27$, $p < .01$). That is, the longer sessions yielded *higher* mean scores on the SAT than the shorter sessions, with the Standard length session showing an increase of 12 points over the Short session, and the Long sessions indicating a 15 point average increase over the Standard session.

In comparison to the CB records, 48% of the participants in the Short session met or exceeded their CB record SAT scores, for the Standard session, 49% of the participants met or exceeded their CB record SAT scores, and for the Long session, 50% of the participants met or exceeded their CB record SAT scores.

Last 50 min Performance Scores. Performance on the set of three tests administered in the last 50 min of each session, and an additional set administered in the penultimate 50 min in the Long condition indicated an interesting pattern of results. For the final 50 min of each session, scores were remarkably similar (Short $M = 34.06$, $sd = 8.85$; Standard $M = 34.48$, $sd = 9.10$; Long $M = 34.34$, $sd = 9.35$). An ANOVA performed on these scores indicated no significant differences across the conditions ($F(2, 366) = .55$, ns; MSE = 5473).

However, when examining the Long session condition, performance in the penultimate 50 min session ($M = 35.42$, $sd = 8.49$) appeared to exceed performance in the final 50 min of either the Short, Standard or the Long conditions. Paired *t*-tests supported this point ($t(188) = 3.25$, $p < .01$, $d = .15$; $t(190) = 2.48$, $p < .05$; $d = .11$; and $t(210) = 3.02$, $p < .01$, $d = .11$, for comparisons with the Short, Standard, and the final 50 min of the Long sessions, respectively.) Although all three comparisons are statistically significant, they are of a small magnitude, according to Cohen's (1988) rules-of-thumb. The mean scores and 95% confidence intervals are shown in **Figure 1**.

As would be expected, given the commonality of test content across both the archival SAT scores and the individual condition SAT equivalence scores (not to mention the part-whole correlations for the last 50 min segments and the overall condition scores), the correlations among all of these measures were consistently high, and are shown in **Table 4**. Of particular importance, however, are the correlations between archival SAT scores and individual condition SAT equivalence scores. These correlations were $r = .87$, .86, and .83 for the Short, Standard, and Long conditions, respectively. These correlations are likely comparable to the test-retest correlations of the SAT under high-stakes testing conditions.

Summary. By keeping to the same overall procedures used in standardized SAT testing (including day of the week, time of day, instructions, and so on), and by including cash bonuses for the participants who met or exceeded their archival SAT records scores, we appear to have created a testing situation that was largely congruent with the typical SAT testing experience, even though the participants were already enrolled in post-secondary institutions at the time they completed the study. Nearly half of the participants obtained scores that met or exceeded their archival SAT scores, even under a condition when an extra hour of testing was involved in a single session. The majority of the remaining participants had 100 or fewer point difference between their archival and in-study SAT scores.

As expected, the conditions of testing (test length) had little effect on the mean SAT equivalence scores. In contrast to those members of the profession or the lay public who would have anticipated a decline in test performance as the test sessions increased in length from 3½ hr to 4½ hr to 5½ hr total time, the results indicated that average SAT equivalence scores actually *increased* with increasing test session length. From the Short to the Standard testing conditions, the increase was 13 points (on the V+ Q test composite), and from the Standard to the Long testing conditions, there was another increase of 15 points. Because this experiment had a within-subjects design with counterbalanced test condition orders, and counterbalanced test forms, these differences cannot be attributed to either sample differences, session order effects or test form differences.

Subjective Measures Taken During Testing.

Subjective Fatigue. Mean levels of subjective fatigue, assessed just prior to the initial SAT test, then repeated after approximately each hour of testing, and finally assessed just after the last section of testing in the Short, Standard, and Long testing conditions are shown in **Figure 2**. Within-condition repeated-measure ANOVAs yielded results that are concordant with the visual representation in the figure. In the Short condition, the test of fatigue over time-on-task was $F(2, 394) = 30.44$, $p < .01$, MSE = .23; in the Standard condition $F(3, 609) = 28.41$, $p < .01$; MSE = .23; and in the Long condition, it was $F(4, 824) = 9.24$, $p < .01$; MSE = .31. That is, with increasing time-on-task, subjective fatigue increased. There were also a few notable comparisons across the conditions. First, pretest subjective fatigue for the Short condition was significantly lower than the pretest subjective fatigue for the Standard condition ($t(185) = -2.71$, $p < .01$), but only marginally lower, in comparison to the Long condition ($t(186) = -1.74$, ns). Immediate posttest assessments of subjective fatigue were also lower for the Short condition, compared to the Standard condition ($t(185) = -3.91$, $p < .01$) and the Long condition ($t(186) = -5.10$, $p < .01$). However, the difference in subjective fatigue at immediate posttest for the Standard condition was not significantly different from the Long condition ($t(187) = -1.61$, ns.). Finally, there was no noticeable decline in subjective fatigue from the last interim administration to the posttest assessment immediately after testing.

Negative Affect. There were substantial positive correlations between negative affect ratings and subjective fatigue (in the aggregate, these were $r = .54$, .60, and .50, all $p < .01$, respectively for the Short, Standard and Long conditions). Given this level of overlap, and the expectations that subjective fatigue and negative affect are similar underlying constructs, it was also expected that negative affect would increase over the course of the time-on-task. The mean levels of negative affect are shown in **Figure 3**. Repeated measures ANOVAs indicated significant effects of time-on-task in the Short condition ($F(2, 396) = 20.37$, $p < .01$, MSE = .09; in the Standard condition $F(3, 606) = 17.78$, $p < .01$; MSE = .09; and in the Long condition $F(4, 816) = 32.40$, $p < .01$; MSE = .11. When compared across conditions, subjective negative affect in the Standard condition posttest was not significantly different from that of the Short condition ($t(186) = -1.24$, ns), though at the 4½ hour mark in the Long condition, negative affect was significantly higher than that in the posttest assessment in the Standard condition ($t(186) = -2.50$, $p < .05$).

Positive Affect. The effects of time-on-task for subjective positive affect had a decidedly different pattern from the subjective fatigue and negative affect measures. In each of the three conditions, positive affect declined during testing, and then largely recovered at the posttest measurement. Mean positive affect ratings are shown in **Figure 4** for the three conditions. Repeated-measures ANOVAs indicated significant differences across the administration times for the Short condition ($F(2, 396) = 5.17$, $p < .01$, MSE = .20; in the Standard condition $F(3, 612) = 11.08$, $p < .01$; MSE = .20; and in the Long condition $F(4, 820) = 7.65$, $p < .01$; MSE = .21. Across conditions, only the posttest positive affect for the Standard condition was significantly higher than the posttest positive affect in the Long condition ($t(187) = 2.04$, $p < .05$).

Positive Motivation for Performance. Subjective reports of positive motivation for test performance showed a pattern of changes over the course of testing relatively similar to that of subjective positive affect, as shown in **Figure 5**. However, no significant differences were found for the Short condition ($F(2, 396) = 2.71$, ns, MSE = .14); though significant differences were found in the Standard condition $F(3, 612) = 4.52$, $p < .01$; MSE = .14; and in the Long condition $F(4, 824) = 9.76$, $p < .01$; MSE = .20. In the Standard and Long conditions, there were significant recoveries of positive motivation from the last assessment during testing to the posttest assessment (($t(206) = -3.69$, $p < .01$; and ($t(207) = -5.33$, $p < .01$, respectively).

Confidence. Subjective confidence for test performance declined during testing in each of the testing conditions, as shown in **Figure 6**. Repeated-measures ANOVAs indicated significant differences across the administration times for the Short condition ($F(2, 396) = 5.02$, $p < .01$, MSE = .19; in the Standard condition $F(3, 612) = 3.74$, $p < .01$; MSE = .22; and in the Long condition $F(4, 820) = 3.70$, $p < .01$; MSE = .23. Subjective confidence appeared to recover slightly at posttest assessments in the Short and Standard conditions, but not in the Long condition. However, at posttest assessment, only the Short and Long conditions were significantly different from one another ($t(186) = 2.72$, $p < .01$).

Summary. Reports of subjective fatigue were relatively low at pretest in each of the three test-length conditions (though slightly elevated in the Standard and Long conditions). Subjective fatigue increases as time-on-test increased in all conditions, though mean levels of subjective fatigue were relatively stable from the last interim questionnaire to the posttest questionnaire. *Clearly, the longer test session conditions resulted in higher mean subjective fatigue reports.* Negative affect also showed a similar pattern of increases with increases in time in each testing session, with the Standard and Long conditions showing higher levels of negative affect than the final interim assessment in the Short condition.

Reports of positive affect started off relatively higher at pretest than during each test, but recovered (generally) to pretest levels at the posttest assessment in each condition, as did reports of positive motivation. Mean confidence showed declines in each condition (though mean confidence levels at pretesting were lower in the Standard and Long conditions than in the Short condition). Recovery of confidence in test performance showed a much shallower increase at posttest than was observed for positive motivation and positive affect.

Across these subjective reports, it appears that participants are, on average, more fatigued as testing time increases. They also report more negative affect, less positive affect, less positive motivation and lower confidence at testing time increases. Longer test times resulted in participants showing greater levels of tiredness and unhappiness, and lower levels of positive affect, in comparison to the shorter testing conditions. At the end of testing, the levels of positive affect show recovery, but subjective fatigue and negative affect do not. How long it takes for the participants to fully recover (in terms of subjective fatigue and negative affect) is not known from this study. (As an aside, when asked about what they were going to do immediately after the testing session, about half the participants said that they were going to eat a meal [54% in the Short condition, 50% in the Standard condition, and 60% in the Long condition], and about one third of the participants said that they were going to take a nap or rest [30% in the Short condition, 34% in the Standard condition, and 31% in the Long condition].)

Self-Efficacy

Both traditional and normative self-efficacy questions were administered with the pretest, interim, and posttest questionnaires. Self-efficacy judgments were highly stable, both within and between sessions. In addition, the traditional and normative methods to measure self-efficacy were also highly correlated. As a result, single composite measures of self-efficacy were created from the 6, 8 and 10 item sets for the Short, Standard, and Long conditions ($\alpha =$ .95, .97, and .98, respectively). Correlations of self-efficacy composites across sessions were also quite high (Short with Standard $r =$ .93, Short with Long $r =$ .90, and Standard with Long $r =$ .93). Self-efficacy composites also correlated substantially with both archival and condition SAT equivalence scores (range $r =$ .60 to .69 for the condition correlations) and range $r =$ .56 to .61 for the correlations between self-efficacy measures and the archival SAT scores. The most interesting finding regarding self-efficacy composites was that there were significant correlations betweens self-efficacy and condition SAT equivalence scores, even after archive SAT scores had been partialled out of the equation. For the Short condition the correlation between self-efficacy and SAT equivalence score (with archive SAT partialled) was $r\ (176) = .22$, $p < .01$. For the Standard condition, the partial correlation was of a similar magnitude $r\ (181) = .26$, $p < .01$. However, in the Long condition, the partial correlation was $r\ (185) = .50$, $p < .01$. Note that a similar pattern of partial correlations was found for subjective Confidence measures (partial r's = .33, .26, and .39 for the Short, Standard and Long conditions, respectively).

What meaning can be inferred from these results is not entirely clear. One possible interpretation is that these self-report measures (self-efficacy and confidence) capture individual differences in proximal motivation that explains variance in performance not accounted for by the SAT performance under high-stakes testing conditions. Another interpretation is that there is sufficient self-generated feedback from the attempting the test problems themselves that indicate to the participant that he/she is performing well or not-so-well in the test. The latter possibility seems less likely in that the pretest self-efficacy measures have a similar pattern to the interim and posttest questionnaires.

Cross-correlations among subjective fatigue measures. The intercorrelations among the subjective fatigue measures in the three test length conditions are shown in **Table 5**. These correlations are all positive and significant, and there are three notable patterns that can be seen from the correlation matrix. First, the repetitions of subjective fatigue assessments within each condition show a simplex-like pattern (e.g., see Ackerman, 1987), with the largest correlations occurring for adjacent administrations (the diagonal entries), and the smallest correlations for those pairs most distant from another in time of administration. This pattern of repeated-measurements correlations is expected, and is ubiquitous to repeated measures of both psychological and physical variables (see Humphreys, 1960 for a discussion). The salient aspect of the within-condition correlations is that there is a high degree of correlation among the subjective estimates, perhaps with the exception of the most distant assessments in the Long condition.

The other two patterns in the correlation matrix pertain to the ‘first’ administration of the subjective fatigue measures in each condition, and the ‘final’ administration of the subjective fatigue measures in each condition. Generally speaking, the highest correlations for the first assessments of subjective fatigue in each condition share more common variance than other cross-correlations with those measures. Similarly, the final administration of subjective fatigue measures in each condition also tend to share the most variance in cross-correlations. That is, even though the final assessment in the Short condition comes at 3.5 hr into the task, it shares more variance with the assessment in the Standard condition (4.5 hr) and the Long (5.5 hr) than with assessments taken at similar times-in-testing. Thirty-eight of the variance on the final Long condition subjective fatigue measure can be accounted for by a measure administered after 2.5 hr in a different condition.

Associations between subjective fatigue and SAT performance. One critical question in this investigation is whether there is a significant and substantial association between individual differences in subjective fatigue during tests of different lengths and actual performance on those tests. Correlations between the subjective fatigue measures and performance on the SAT are shown in **Table 6**. Specifically provided are correlations between the pretest, interim, and posttest subjective fatigue measures (along with a composite measure) and the SAT taken under high-stakes conditions, the overall and last 50 min SAT performance during which the subjective fatigue measures were administered. (In the Long testing condition, also shown are performance data from the next to last 50 min module of the SAT test).

First, it should be noted that the none of the subjective fatigue measures were significantly correlated with archival SAT performance. In contrast, there were some significant negative correlations between subjective fatigue during testing and the SAT performance (both overall and during the last 50 min of testing), meaning that higher levels of subjective fatigue were associated with lower SAT performance. In the Short and Standard length conditions, none of these correlations exceeded a magnitude of .20. However, in the Long test length condition, some of the correlations were more substantial. The largest correlations (r = -.29) were found between the Interim 1 assessment of subjective fatigue (which occurred at 2½ hr into the session) and both overall SAT performance and performance in the last 50 min of testing. While these correlations are not large in magnitude, they do point to an association between subjective feelings of fatigue and performance on the SAT test, at least in the Long test length condition.

Trait Complexes

The distal trait measures (except for self-concept/self-estimates of ability) were subjected to factor analysis in order to develop trait complex measures. After the first iteration (where a few scales with very low communalities were eliminated), the final set of 42 measures was subjected to factor analysis. Six broad trait complex factors were derived and rotated to a varimax criterion. Unit-weighted z-score composites were formed from the scales that had salient loadings on the factors, resulting in trait complex scores. Trait complexes were identified as I. Need for Achievement/Mastery (nAch/Mastery), II. Desire to Learn/Typical Intellectual Engagement (DTL/TIE); III. Neuroticism/Anxiety (N/Anxiety); IV. Extroversion; V. Extrinsic Goals, and VI. Competitiveness/Other-Oriented Goals. Given the relatively high degree of bandwidth for these trait complexes, high levels of internal consistency were not expected. Nonetheless, these trait complex composite scores had reasonable levels of α, ranging from .70 to .89, as shown in **Table 7**. Correlations among the trait complex composites were generally smaller than r = .30, except for the correlation between I and II, which was r = .44. The correlations among the trait complex scores are also shown in the table.

Self-Concept and Self-Estimates of Ability

Scales of self-concept and self-estimates of ability for math and verbal domains were computed. Because the same-domain scales (math self concept and math self-estimate of ability, and verbal self concept and verbal self-estimates of ability) were relatively highly correlated, r = .76 for math, and .70 for verbal (e.g., see Ackerman & Wolman, 2007), composites for math self-concept/self-estimates of ability and verbal self-concept/self-estimates of ability were created. As expected from prior literature, the verbal and math composites were essentially unrelated to one another (r = .06), even though objective math and verbal abilities tend to be correlated around r = .40.

Correlations between Trait Complexes, Self-Concept/Self-Estimates of Ability and SAT Performance.

Correlations between trait complex scores, self-concept/self-estimates of ability and SAT scores (archival, SAT equivalence scores from each condition, and last the 50-min segments) were computed, and are shown in **Table 8.** On the one hand, for the trait complex scores, only the DTL/TIE trait complex showed consistently significant correlations, albeit of a relatively modest magnitude (r = .15 to .23). On the other hand, self-concept/self-estimates of ability composites, especially for the math domain, showed more robust correlations with SAT scores. The math self-concept/self-estimates correlated r = .38 with archival SAT scores, and from $r = .30$ to .44 for the SAT scores obtained in the current testing sessions. Of particular importance, however, is the general lack of salient correlations between the Neuroticism/Anxiety trait complex and SAT performance. Although a few correlations were significant, the magnitude was less than $r = .20$ across the conditions.

Correlations between Trait Complexes, Self-Concept/Self-Estimates of Ability and Self-Report measures during SAT testing.

In contrast to the correlations with SAT test scores, the trait complexes showed a robust pattern of significant and salient correlations with the self-report measures obtained before, during and after the SAT testing. **Table 9** provides the correlations for each of the administrations of subjective fatigue, negative affect, positive affect, positive motivation, confidence, and self-efficacy. In particular, Subjective Fatigue showed positive correlations with the Neuroticism/Anxiety trait complex, and negative correlations with both the nAch/Mastery and DTL/TIE. Negative affect showed significant positive correlations with Neuroticism/Anxiety throughout the test sessions, while Positive Affect and Positive Motivation showed positive correlations with nAch/Mastery and DTL/TIE. Positive affect also showed small, but occasionally significant negative correlations with Neuroticism/Anxiety and positive correlations with Math self-concept/self-estimates of ability. Confidence measures shared the most communality with trait complexes and self-concept/self-estimates of abilities, correlating positively with nAch/Mastery, DTL/TIE, Math and Verbal self-concept/self-estimates of ability, and negative correlations with Neuroticism/Anxiety. Self-efficacy measures were most highly positively associated with Math and Verbal self-concept/self-estimates of ability, DTL/TIE, and negatively associated with Neuroticism/Anxiety.

Accounting for Individual Differences in Subjective Fatigue. Even though individual trait complexes (especially nAch/Mastery, DTL/TIE, and N/Anxiety) and, to a lesser degree, self-concept/self-estimates of ability were individually significantly correlated with the measures of subjective fatigue, overall prediction must take account of the covariance among these predictor variables. The results of two sets of hierarchical regressions for the pretest and posttest measures of subjective fatigue are provided in **Table 10**. For the first model, the regressions indicate that trait complexes and self-concept/self-estimates of ability from the at-home questionnaire account for from 22 to 30% of pretest subjective fatigue, and from 22 to 25% of the variance in posttest fatigue across the three testing conditions. For prediction of posttest fatigue, adding the level of pretest fatigue to the regression resulted in significant and substantial increases in variance accounted for (by 14 to 30% variance accounted for), yielding final amounts of variance accounted for of 55%, 53% and 36% for the Short, Standard, and Long testing conditions, respectively.

If we consider posttest fatigue measured in the Short testing condition as a potential predictor of final posttest fatigue in the Standard and Long testing conditions (Model 2), it is clear that individual differences in the level of subjective fatigue experienced during a relatively short testing session, accounts for a substantial degree of variance in subjective fatigue in the longer testing sessions, both at initial pretest assessment and at final posttest assessment. Including Short condition posttest subjective fatigue resulted in increases of variance accounted for at pretest 15% and 11%, respectively, for the Standard and Long conditions, and increased variance accounted for at posttest 33% and 24% respectively, for the Standard and Long conditions. For posttest fatigue, including pretest fatigue from the same condition also provided significant incremental predictive validity of 9% and 4% respectively, for the Standard and Long conditions. In the final models for posttest fatigue in the Standard and Long conditions, the aggregate predictions from trait complexes/self-concept/self-estimates of ability, Short condition posttest fatigue, and pretest fatigue, resulted in accounting for 66% of the variance in the Standard testing condition, and 50% of the variance in the Long testing condition.

In summary, roughly 20-30% of the variance in pretest and posttest subjective fatigue is predictable from trait measures assessed at a time different from the actual testing day. An additional 11-15% of the variance in pretest fatigue is accounted for by posttest subjective fatigue in a Short SAT testing session (or in contrast, 14-30% of the variance is accounted for by pretest fatigue on the testing day). If the posttest fatigue from the Short testing session is included in the prediction equation, the amount of variance accounted for by pretest in-session subjective fatigue for posttest fatigue is still significant, but is reduced to 4-9% of the variance.

There were relatively minor or non-significant correlations between actual SAT test performance in the last 50 min of task performance and subjective fatigue immediately after test performance (raw correlations were $r = -.16$, $p < .05$; = .09, ns, and -.13 ns for the Short, Standard and Long conditions). However, as a final step in the regressions, we added actual SAT performance in the last 50 min of testing in the Short, Standard, and Long conditions to the prediction of post-task subjective fatigue. In each of the analyses, individual differences in objective SAT performance did not account for a significant amount of variance. Incremental predictive validity of the SAT performance in predicting subjective fatigue was either 0 or 1% of variance accounted for.

Incremental predictive validity of subjective fatigue on SAT performance

When the associations of subjective fatigue and SAT performance are considered in the context of distal predictors of performance (namely the trait complexes and self-concept/self-estimates of ability), it is clear that subjective fatigue does add significant incremental predictive validity, but only in the Long condition (see **Table 11**). For the final 50 min of SAT test performance in the Long testing condition, subjective fatigue accounted for 4% of the variance, after the trait complexes and self-concept/self-estimates of ability have been entered into the prediction equation. For the next to last 50 min of SAT performance in the Long testing condition, subjective fatigue accounted for 6% of the variance, after the distal predictors were entered. In all conditions, trait complex scores and self-concept/self-estimates of ability account for approximately 25% of the variance in SAT performance during the last 50 min of testing. Beyond this predictive power, subjective fatigue measures do not provide incremental predictive validity for SAT test performance in either the Short and Standard conditions, and only account for an additional 4-6% of the variance in the Long testing condition.

Accounting for Aggregated Subjective Reports. Given the high correlations among each of the measures of subjective fatigue, negative affect, positive affect, positive motivation, and confidence across the three sessions, composites were created to allow for aggregation to help possibly illuminate the overall relations between the different constructs. **Table 12** shows the correlations between self-report measures aggregated across multiple samples during each condition. The cross-condition correlations show that 35-50% of the variance in subjective fatigue was common, and similar levels of common variance were found for the other measures. The aggregated measures also show consistent patterns of correlations with the trait complex composite scores and self-concept/self-estimates of ability (see **Table 13**), and for correlations with SAT performance (see **Table 14**). A final aggregation (across Short, Standard, and Long conditions) was obtained. The correlations among these aggregated measures are shown in **Table 15**. These correlations clearly indicate the positive overall relationships between subjective fatigue and negative affect, and the negative correlations between subjective fatigue and confidence.

Hierarchical regressions were then performed for each of the composite measures, using Trait Complexes, Self-Concept/Self-Estimates of Ability, and SAT performance under high-stakes testing as predictor variables. The results of these analyses are provided in **Table 16**. Across the three experimental sessions, 31% of the variance in subjective fatigue was accounted for by the trait complexes, but no significant incremental predictive validity was found for self-concept/self-estimates of ability or prior SAT scores. For aggregated negative affect, 18% of the variance was accounted for by the trait complexes, and prior SAT performance accounted for an additional 4% of the variance. The amount of variance in positive affect accounted for by the trait complexes was 17%, and self-concept/self-estimates of ability accounted for an additional 4%. Positive motivation was not significantly accounted for by any of the predictor variables. Finally, confidence for performance on the SAT was well accounted for by the trait complexes (36% of variance accounted for) and self-concept/self-estimates of abilities (7% additional variance accounted for). Prior SAT performance accounted for a small, but significant additional 2% of the variance in confidence.

Summary. In attempting to determine what variables account for the variance in subjective fatigue and other self-report measures (across the conditions), it is clear that the trait complexes have the largest source of common variance, accounting for 31% of the variance in subjective fatigue, 36% of the variance in confidence, and lesser amounts for negative affect and positive affect. Self-concept/self-estimates of abilities account for small, but significant incremental validity for positive affect and confidence reports. Individual differences in performance on the SAT under high-stakes testing conditions do not account for any incremental variance in subjective fatigue, but do account for small amounts of variance in negative affect and confidence.

Sleep and Caffeine Consumption

For amount of sleep the night before testing, no appreciable correlations were found with SAT scores overall for the last 50 min of testing. However, amount of sleep did show negative correlations with composite measures of subjective fatigue ($r =$ -.19, -.25, and -.28 for the Short, Standard, and Long conditions, respectively, all $p <$ .01), no significant correlations with negative affect, and positive correlations with positive affect ($r =$ .25, .20, and .24 for the Short, Standard, and Long conditions, respectively, all $p <$ .01). Thus, participants who had more sleep the night before testing were somewhat less likely to feel fatigued, and somewhat more likely to express positive affect during testing.

Individuals who reported more dependence on caffeine were more likely to report subjective fatigue before, during, and after the SAT testing. Across the testing sessions, the correlations between caffeine consumption and subjective fatigue were $r =$ .22, .25, and .18, all $p <$ .01, for the Short, Standard and Long conditions, respectively. The only significant correlation with SAT test performance was for the Short condition SAT equivalence score (r = -.15).

SAT Attitudes and Experiences

Although interpretation of single item questions should be done cautiously, there were a few notable results with the additional questions in the AHQ, with respect to subjective fatigue and SAT performance. Individuals who agreed with a statement "I felt stressed when I took the SAT" had higher levels of subjective fatigue in all conditions of the experiment (r = .25, .23, and .33, for the Short, Standard, and Long conditions), and lower levels of archival SAT performance ($r =$ -.29), all $p <$ .01. A similar pattern of results were found for individuals who agrred with the statement "I think my anxiety negatively influenced my perofrmance on the SAT" ($r =$ .23, .31, .28, with subjective fatigue) and $r =$ -.39 with SAT performance. The number of times an examinee took the SAT during high school had nonsignificant relations with subjective fatigue and archival SAT performance. The number of practice tests completed by the examinee had only modest or nonsignificant relations with both subjective fatigue ($r =$ .14, ns; .13 ns; and .17, $p <$ .05, for Short, Standard and Long conditions, respectively) and archival SAT performance ($r =$ .15, $p <$ *.05)*. For this sample of participants, 38% reported having taken some SAT preparation course. There was no significant relationshipw between the SAT preparation course experience and subjective fatigue in any of the test conditions.

Post-Session Questionnaires

At the end of each session, participants were asked a set of questions about their overall session experience and strategy. In response to a question about effort levels (selecting from increasing effort during the session, keeping a constant level of effort, decreasing effort, or increasing, then decreasing effort), in the Short testing condition, 16% of the participants reported increased effort, 61% reported a constant level, 13% reported a decrease in effort, and 10% reported increasing, then decreasing effort. In the Standard condition, the rates were 18%, 54%, 14%, and 14%, respectively. In the Long condition, the rates were 16%, 50%, 18%, and 16%, respectively. Thus, a relatively constant proportion of participants increased effort during the three sessions, with small decreases in the proportion of those participants reporting a constant level of effort, and a concomitant increase in those participants who reported some decrease in effort in the longer conditions.

An analysis of last 50 minute SAT test performance by responses showed significant differences among these groups in the Short condition ($F\,(3,199) = 3.54$, $p < .05$), with the highest scores obtained by the participants who reported a constant level of effort (M = 35.15 items) and the lowest scores obtained by the participants who reported a reduction of effort (M = 29.16). No significant differences were found in either the Standard or Long condition between participants who reported increasing, constant, or decreasing levels of efforts across the session.

For performance effects, **Figure 7** shows a breakdown of last 50 min SAT performance scores by responses to the effort strategy question. Individuals who reported constant effort levels across the three conditions had consistently higher cores than those who reported decreased effort during testing. Those that reported an increase then a decrease in effort only performed relatively well in the Standard condition, in comparison to those who reported constant levels of effort. The differences in performance on the last 50-min segments between constant and decreased effort were approximately 6 questions for the Short condition, 4.5 questions for the Standard condition, and 3.3 questions for the Long condition.

For comparison purposes, the difference between the constant and decreasing effort groups on the archival SAT data were relatively small (in the Short condition, SAT = 1255 vs. 1228; in the Standard Condition, SAT = 1250 vs. 1227; and in the Long condition SAT = 1265 vs. 1233, respectively). However, for the SAT equivalence scores for the three conditions, the differences were more substantial, especially in the Short and Long conditions (in the Short condition SAT = 1230 vs. 1140; in the Standard condition SAT = 1237 vs. 1196; and in the Long condition SAT = 1252 vs. 1180). Similarly those individuals who reported a decrease in effort or an increase then a decrease in effort tended to also report greater levels of subjective fatigue throughout each condition, as shown in **Figure 8**.

Participants were also asked to compare the SAT test experience to that of completing a course final exam. Options were "less fatiguing than taking a course final exam" "about the same level of fatigue as taking a course final exam" or "more fatiguing than taking a final course exam." In the Short condition, 37% reported the test was less fatiguing, 19% reported about the same, and 44% reported that the SAT was more fatiguing than a final exam. The difference in performance between these groups of individuals was marked and significant (F (2, 200) = 10.44, $p <$ .01). The mean levels of performance for the last 50 min of SAT testing was 36.8, 34.7, and 31.0, respectively.

In the Standard condition, the response rates were 31%, 20%, and 50% respectively, with a similar pattern of performance on the last 50 min of the SAT (F (2, 203) = 11.24, $p <$ .01), M = 38.0, 34.8, and 31.5 respectively. In the Long condition, only 20% of the participants reported the session was less fatiguing than a final exam, 17% reported it to be about the same, and 63% reported it was more fatiguing than a final exam. Performance differences among these groups were attenuated (M = 36.7, 36.8, and 33.2), and the differences were not significant (F (2, 205) = 2.94, p = .05).

Post-session questionnaire (in the third session only)

At the end of the third session (after all of the testing had been completed), participants were asked to agree or disagree with two key questions. The first question ("My mood influenced how well I performed on the SAT") correlated significantly negatively with SAT performance under high-stakes conditions (r = -.22, $p <$ .01) and in the three experimental conditions (r = -.27, -.25, and -.27 for the Short, Standard and Long conditions, respectively). In other words, those individuals who believed that their mood influenced their SAT performance, tended to obtain somewhat lower scores on the SAT across the various conditions. Responses to this question also significantly correlated with aggregate scores on the subjective ratings during the experiment, especially Subjective Fatigue (r = .32) and Confidence (r = -.42). That is, those individuals believed that their mood influenced their SAT performance also had higher levels of subjective fatigue, and lower levels of confidence in their performance during SAT testing.

The other question ("It was hard to keep up my motivation on the SAT") had a somewhat smaller, but still significant correlation with SAT performance under high-stakes conditions (r = -.16), and significant correlations with SAT performance in the three experimental conditions (r = -.22, -.25, and -.26 for the Short, Standard and Long conditions, respectively). That is, those individuals who stated that it was difficult to keep up their motivation on the SAT, tended to obtain somewhat lower scores on the SAT across the various conditions. Responses to this question also significantly and substantially correlated with aggregate scores on the subjective ratings during the experiment, especially Subjective Fatigue (r = .50) and Confidence (r = -.50). Those individuals found it difficult to keep up their motivation, experienced substantially higher had levels of subjective fatigue during the study, and substantially lower levels of confidence in their performance during SAT testing. In addition, this question correlated positively with the aggregate measure of negative affect (r = .24, $p <$ .01), and negatively with positive affect (r = -.34) and positive motivation (r = -.30).

When compared with the effort strategy question administered at the end of each session, there was a clear association between those individuals who found it hard to keep up motivation on the SAT, and those that reported decreasing effort during the course of the testing session (see **Figure 9**). For the 44 participants who reported decreasing effort in the Short condition, mean response on the 'motivation difficulty' item was $M = 4.61$; for the 140 participants who reported stable or increasing effort in the Short condition, mean response was $M = 3.81$ ($F(1,182) = 9.15$, $p < .01$). Similar results were obtained for the Standard length condition $M = 4.85$, $M = 3.68$ respectively, for the 48 participants who reported reduced effort during the test and the 132 participants who reported stable or increasing effort) $F(1,178 = 22.16$, $p < .01$), and for the Long test condition ($M = 4.53$, $M = 3.74$) for the 60 participants who reported reduced effort during the test and the 121 participants who reported stable or increasing effort; $F(1,179) = 10.97$, $p < .01$. Although these associations are not large in magnitude, they present a consistent picture of individuals who were experiencing subjective fatigue during testing, a reaction that indicated a reduction in effort on the test, along with a reduction in confidence in performance.

Discussion

SAT Performance Under Conditions of different Test Lengths. Perhaps the most salient results from this study were found in the SAT equivalence scores for the three different testing conditions. Scores in the Short test condition were not higher than scores in the Standard or Long conditions. To the contrary, scores in the Standard and Long conditions were higher than scores in the Short condition (the mean score for the Standard condition was 13 points higher than the mean score in the Short condition, and the mean score for the Long condition were another 14 points higher than the Standard condition). If anything, the longer testing times lead to higher levels of performance, on average.

Mean Levels of Subjective Fatigue. Similar to many findings in the historical literature on subjective fatigue during cognitively intensive tasks (see Ackerman & Kanfer, 2006, for a review), there was, at the level of means, a dissociation between levels of subjective fatigue and actual test performance. That is, across the three sessions, mean subjective fatigue increased with additional time-on-task, by significant amounts. At the end of testing, mean subjective fatigue was higher in the Standard condition than the Short condition, and mean subjective fatigue was even higher in the Long testing condition. Clearly, spending as little as a few hours of cognitive ability testing yields marked increases in subjective reports of fatigue. However, these subjective reports did not correspond to a concomitant decrease in average cognitive test performance in any of the testing conditions.

Individual Differences in Subjective Fatigue

The first notable result with respect to individual differences in subjective fatigue was that the assessments taken *prior* to the start of testing accounted for a large amount of the variance in subjective fatigue measures taken at the end of testing, especially in the Short and

Standard length testing conditions. Pretest fatigue accounted for 49% of the variance in posttest fatigue in the Short testing condition, and pretest subjective fatigue accounted for 46% of the variance in posttest subjective fatigue in the Standard length condition. The amount of variance accounted for in the Long test length condition was somewhat less in comparison, only 23% of the variance in posttest subjective fatigue was accounted for by pretest subjective fatigue.

Distal personality/interest/motivation trait complexes accounted for significant levels of subjective fatigue across all three conditions, and throughout the testing session. The most salient correlations were found with Neuroticism/Anxiety (positive), Desire to Learn/Typical Intellectual Engagement (negative), and nAch/Mastery (negative). That is, individuals who were more likely to report higher levels of general neuroticism and anxiety were also more likely to report higher levels of subjective fatigue throughout the study, while individuals who had high levels of DTL/TIE and nAch/Mastery reported lower levels of subjective fatigue throughout the study. In the aggregate, distal trait complex predictors accounted for a range from 22 to 30 percent of the variance in subjective fatigue measures. Self-reports of pretest subjective fatigue accounted for an additional 14 to 30 percent of variance in posttest subjective fatigue, meaning that from 36 to 55 percent of the variance in posttest subjective fatigue could be accounted for by measures administered *prior* to any actual performance engagement with the SAT test.

With respect to the relationship between subjective fatigue and test performance, when the testing time was extended to a 5½ hr session (the Long condition), subjective fatigue was a significant predictor of test performance, after distal predictors of personality, interest, self-concept, and so on, were entered into the equation. The amount of test performance variance accounted for by subjective fatigue under such conditions was less than 10% in isolation, and 4-6% after taking the distal predictors into account.

Conclusions

This study represents another instance of the paradox that has been associated with the construct of cognitive fatigue over the past Century. That is, even when the SAT situation was extended from 3½ to 4½ to even 5½ hr of nearly continuous testing in a single sitting, mean levels of performance not only did not decline, they actually increased, on average (though by relatively small amounts). However, as testing progressed within each condition and between conditions of different testing times, self reports of fatigue showed marked increases. Participants also reported higher levels of negative affect, lower levels of positive affect, positive motivation, and confidence about their performance on the SAT tests. At the aggregate level of performance and self-reports, there is a clear pattern of stable (or slightly increased) cognitive performance and increasing subjective fatigue as testing proceeded over increasingly longer periods of time.

Although mean subjective fatigue increased with increasing time-on-task, individual differences in subjective fatigue were at least partly accounted for by individual differences in stable traits. About 30% of the variance in subjective fatigue across the three testing conditions was accounted for by trait complex scores, suggesting that a substantial portion of the variance in self-reports of subjective fatigue is likely to be attributable to stable dispositions. Whether such reports are transituational or operate only in the context of cognitive ability testing is an open question. Nonetheless, individual differences in posttest subjective fatigue in the Short testing condition (3½ hr of testing), also accounted for a significant and substantial amount of variance in posttest fatigue in the Standard and Long conditions, indicating that it may be that just the "fact" of testing also contributes markedly to self-reports of fatigue after longer testing times. Individual differences in actual SAT performance contributed no significant variance to predicting subjective fatigue after the stable trait measures had been entered into the equation for the Short and Standard conditions, suggesting that, ceteris paribus, individuals who are performing poorly on the SAT are *not* more likely to be more fatigued than individuals who performed better on the SAT.

In the final analysis, longer testing times did lead to increases in reports of subjective fatigue, that did not recover immediately at the end of the testing session. It might be reasonably expected that longer testing times in an operational environment will lead to increased complaints from test takers along these lines. However, much of the variance in self-reported fatigue appears to be predictable from stable traits. In the Standard test length condition, the amount of variance in subjective fatigue accounted for by the treatment (i.e., time-on-task from pretest to posttest) was only 4% ($d = .39$; see Cohen, 1988). In the Standard condition, the amount of posttest subjective fatigue accounted for by pretest subjective fatigue was 47%. In comparison, trait complexes accounted for 27% of the variance in pretest subjective fatigue, and 24% variance in posttest subjective fatigue.

From the perspective of the examinee, this research provides a somewhat complex message. On the one hand, the results suggest that examinees will quite likely experience subjective fatigue during testing, especially as the time-on-task proceeds. The experience of fatigue does not appear to be, in and of itself, detrimental to performance on the test. Individuals who reported subjective fatigue during testing did not appear to perform markedly worse on the test than individuals who did not feel fatigued. On the other hand, there did appear to be a difference in performance, especially in the last hour, between those individuals who decreased their effort on the test and those who either maintained constant effort or increased effort over the course of each session.

Ultimately, subjective reports of subjective fatigue appear to be mostly determined by the individual's personality traits and the 'fact' of the test context (which is instantiated prior to the individual's immediate contact with the test). Test preparation experiences or number of practice tests completed by the participants did not appear to mitigate these influences. Perhaps there may be some beneficial effects if individuals who are at-risk for high levels of subjective fatigue are told that this is to be expected, and that feelings of fatigue are not impeding to performance unless the individual reduces effort in the task. Whether or not providing such information or some kind of intervention (such as relaxation or similar stress-reduction techniques) would have effects on subjective fatigue, is an open question that would require a separate empirical investigation.

References

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102*, 3-27.

Ackerman, P. L., & Beier, M. E. (2006). Determinants of domain knowledge and independent study learning in an adult sample. *Journal of Educational Psychology, 98*, 366-381.

Ackerman, P. L., Bowen, K. R., Beier, M. B., & Kanfer, R. (2001). Determinants of individual differences and gender differences in knowledge. *Journal of Educational Psychology, 93*, 797-825.

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin, 121,* 219-245.

Ackerman, P. L., & Kanfer, R. (2006). *Test length and cognitive fatigue*. Final report to The College Board. Author.

Ackerman, P. L., & Wolman, S. D. (2007). Determinants and validity of self-estimates of abilities and self-concept measures. *Journal of Experimental Psychology: Applied*, *13*, 57-78.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology, 67*, 319-333.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Costa, P. T., Jr. & McCrae, R. R. (1992). *Revised NEO Personality Inventory and Five-Factor Inventory professional manual*. Odessa, FL: Psychological Assessment Resources.

Ebbinghaus, H. (1896-97). Über eine neue Methode zur Prüfung geistiger Fähigkeiten und ihre Anwendung bei Schulkindern. [On a new method for testing mental abilities and its use with school children.] *Zeitschrift für Psychologie und Psysiologie der Sinnesorgane, 13*, 401-459. (Trans. by Wilhelm, 1999).

Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology, 43*, 522-527.

Goff, M., & Ackerman, P. L. (1992). Personality-intelligence relations: Assessing typical intellectual engagement. *Journal of Educational Psychology, 84,* 537-552.

Goldberg, L. R. (2005). *International Personality Item Pool.* http://ipip.ori.org.

Gray, J. A. (1990). Brain systems that mediate both emotion and cognition. *Cognition & Emotion, 4*, 269-288.

Heggestad, E., & Kanfer, R. (2000). Individual differences in trait motivation: Development of the Motivational Trait Questionnaire (MTQ). *International Journal of Educational Research, 33,* 751-776.

Holland, J. L. (1973). *Making vocational choices: a theory of careers*. Englewood Cliffs, NJ: Prentice Hall.

Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology, 4*, 97-110.

Humphreys, L. G. (1960). Investigations of the simplex. *Psychometrika*, *25*, 313-323.

Kanfer, R., & Ackerman, P. L. (2000). Individual differences in work motivation: Further explorations of a trait framework. *Applied Psychology: An International Review, 49,* 470-482.

Kanfer, R. & Heggestad, E. (1997). Motivational traits and skills: A person-centered approach to work motivation. In L. L. Cummings and B. M. Staw (Eds.), *Research in Organizational Behavior* (Vol. 19, pp. 1-57). JAI Press, Greenwich, CT.

Lamb, R. R., & Prediger, D. J. (1981). *The Unisex Edition of the ACT Interest Inventory*. Iowa City, IA: American College Testing.

McNair, D. M., Lorr, J., & Droppleman, L. F. (2003). *Profile of Mood States™ Standard Form.* North Tonawanda, NY: Multi-Health Systems.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement, 53,* 801 - 813.

Rolfhus, E. L., & Ackerman, P. L. (1996). Self-report knowledge: At the crossroads of ability, interest, and personality. *Journal of Educational Psychology, 88*, 174-188.

Spielberger, C. D. (1983). *The State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Viswanathan, M. (1993). Measurement of individual differences in preference for numerical information. *Journal of Applied Psychology, 78*, 741-752.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology, 54,* 1063-1070.

Table 1. *Trait measures. Number of items, mean, sd, and internal consistency reliability.*

Scale		#items	Mean	*sd*	α
1.	Extroversion (IPIP)	15	58.78	11.52	.90
2.	need for Achievement	10	45.47	6.67	.85
3.	Cautiousness	7	29.00	5.43	.82
4.	Impulse Control	7	29.07	5.71	.78
5.	Self-Discipline	9	30.74	7.29	.85
6.	TIE	59	230.74	31.92	.92
7.	Numerical Preferences	11	48.56	10.71	.92
8.	Desire to Learn	8	35.13	5.90	.81
9.	Mastery	8	35.41	5.90	.82
10.	Other-oriented goals	7	27.77	6.09	.85
11.	Competitive Excellence	6	21.90	6.28	.86
12.	Worry	10	36.88	8.15	.84
13.	Emotionality	9	28.89	7.28	.78
14.	Test Anxiety	5	16.55	4.92	.74
15.	Task Value	6	26.77	4.40	.77
16.	Self-efficacy for Learning and Performance	8	35.25	5.97	.88
17.	Extrinsic Goals	4	19.03	3.40	.67
18.	Intrinsic Goals	4	16.29	3.35	.69
19.	Control of Learning Beliefs	4	18.85	2.80	.65
20.	Peer Learning	3	10.13	3.07	.63
21.	Help seeking	4	14.33	3.79	.62
22.	Metacognitive Self-Regulation	12	45.29	7.43	.73
23.	Time and Study Environmental Management	8	31.12	5.75	.69
24.	Effort Regulation	4	16.30	3.38	.64
25.	Critical Thinking	5	17.28	4.36	.79
26.	Organization	4	14.57	3.94	.66
27.	Elaboration	6	24.70	4.29	.66
28.	Rehearsal	4	15.95	3.49	.59
29.	Trait Anxiety (STAI)	20	57.49	12.52	.89
30.	Investigative Interests	15	54.56	14.52	.90
31.	Artistic Interests	15	54.93	15.43	.90
32.	Conventional Interests	15	46.17	15.62	.93
33.	Realistic Interests	15	49.59	12.61	.87
34.	Neuroticism	12	38.37	9.83	.86
35.	Extroversion (NEO-FFI)	12	47.54	8.07	.79
36.	Openness to Experience	12	46.92	6.96	.63
37.	Agreeableness	12	48.40	7.56	.76
38.	Conscientiousness	12	49.71	8.22	.84

Table 1 continued. *Trait measures. Number of items, mean, sd, and internal consistency reliability.*

	Scale	#items	Mean	*sd*	α
39.	Self-Consciousness Focus	5	21.47	3.86	.71
40.	BIS (Behavioral Inhibition System)	7	3.98	.78	.79
41.	BAS Reward Responsiveness	5	4.90	.66	.76
42.	BAS Drive	4	3.99	.79	.73
43.	BAS Fun Seeking	4	4.28	.86	.72
44.	Caffeine Consumption	7	2.93	1.17	.84
45.	Morningness/Eveningness	19	42.31	9.23	.81
46.	Verbal Self Concept	6	4.94	.77	.87
47.	Math Self Concept	6	4.96	.98	.92
48.	Self-Estimate of Verbal Ability	3	74.81	15.72	.89
49.	Self-Estimate of Math Ability	5	76.74	16.14	.92
50.	Self-Estimate of Spatial Ability	3	73.24	16.11	.82
51.	Self-Estimate Memory Ability	6	71.08	16.56	.91
52.	Self-Estimate of General ability	1	81.50	13.12	--

Notes: TIE = Typical Intellectual Engagement; BAS = Behavioral Activation System; IPIP = International Personality Item Pool; NEO-FFI = NEO Five-Factor Inventory; STAI = State-Trait Anxiety Inventory

Table 2. *Layout of Test Sections by Condition*

Time (hours: min:sec)

	Long Condition	**time**	**cumulative**	**Standard Condition**	**Short Condition**	
1.	Pretest questionnaire	5:38	5:38	Pretest questionnaire	Pretest questionnaire	
2.	General and Essay instructions	3:54	9:32	General and Essay instructions	General and Essay instructions	
3.	Essay test (25 min)	25:39	35:11	Essay test	Essay test	
4.	Verbal Test (25 min)	25:48	1:00:59	Verbal Test	Verbal Test	
5.	Break (5 min)	5:31	1:06:30	Break	Break	
6.	Math (25 min)	25:38	1:32:08	Math	Math	
7.	Verbal (25 min) + 1min "stretch"	26:39	1:58:47	Verbal	Verbal	
8.	Math (25 min)	25:25	2:24:12	Math	Math	
9.	Break (5 min)	5:25	2:29:37	Break	Break	
10.	Interim 1 Questionnaire	4:28	2:34:05	Interim Questionnaire	Interim Questionnaire	
11.	Writing (25 min)	25:25	2:59:30	Writing	*Verbal (20 min)*	20:24
12.	Math (25 min)	25:24	3:24:54	Math	*Math (20 min)*	20:25
13.	Break (5 min)	5:25	3:30:19	Break	*Writing (10 min)*	10:23
14.	Interim 2 Questionnaire	4:28	3:34:47	Interim Questionnaire	Posttest Questionnaire	5:00
15.	*Verbal (20 min)*	20:24	3:55:11	***Verbal (20 min)***	**[Total Time: 3:30:58]**	
16.	*Math (20 min)*	20:25	4:15:36	***Math (20 min)***		
17.	*Writing (10 min)*	10:23	4:25:59	***Writing (10 min)***		
18.	Break (5 min)	5:25	4:31:24	Posttest Questionnaire		
19.	Interim 3 Questionnaire	4:25	4:35:49	**[Total Time: 4:31:06]**		
20.	*Verbal (20 min)*	20:23	4:56:12			
21.	*Math (20 min)*	20:23	5:16:36			
22.	*Writing (10 min)*	10:30	5:27:07			
23.	Posttest Questionnaire	5:00	5:33:07			

[Total Time: 5:33:07]

Note: Common 50 minute SAT test segments shown in italics.

Table 3. *Pretest, Interim, and Posttest Self-Report measures. Number of items and internal consistency reliability (from short condition pretest administration).*

Pretest, Interim, and Posttest Scales

Scale	#items	α
1. Subjective Fatigue	12	.91
2. Negative Affect	12	.87
3. Positive Affect	7	.89
4. Positive Motivation	5	.88
5. Confidence	7	.87

Table 4. *Correlations among SAT Records Composite and SAT test performance in current study*

	1	2	3	4	5	6	7
1. CB Records SAT (V+ Q)							
2. Short Session SAT	.87						
3. Standard Session SAT	.86	.86					
4. Long Session SAT	.83	.87	.86				
5. Short -- last 50 min	.77	.86	.79	.82			
6. Standard -- last 50 min	.84	.86	.86	.85	.78		
7. Long - next to last 50 min	.80	.82	.81	.88	.80	.82	
8. Long - last 50 min	.81	.85	.82	.86	.83	.84	.86

df (range from 189 to 211). All correlations significant, $p < .01$

Table 5. *Cross-correlations among subjective fatigue measures.*

	1	2	3	4	5	6	7	8	9	10	11
1. Short Pretest											
2. Short Interim	**.74**										
3. Short Posttest	**.70**	**.86**									
4. Standard Pretest	***.68***	.54	.57								
5. Standard Interim 1	.48	.57	.60	**.74**							
6. Standard Interim 2	.54	.60	.69	**.72**	**.86**						
7. Standard Poshest	.52	.62	**.72**	**.68**	**.80**	**.90**					
8. Long Pretest	***.53***	.46	.46	***.58***	.52	.53	.49				
9. Long Interim 1	.37	.52	.57	.44	.58	.59	.55	**.67**			
10. Long Interim 2	.36	.53	.61	.45	.60	.64	.64	**.61**	**.83**		
11. Long Interim 3	.34	.47	.58	.41	.58	.61	.65	**.47**	**.69**	**.87**	
12. Long Posttest	.35	.47	**.62**	.38	.55	.64	**.69**	**.48**	**.66**	**.82**	**.90**

Note: Boldface = within condition correlations. Italics = Correlations between pretest measures; Gothic type = correlations between posttest measures

Table 6. *Correlations between subjective fatigue and SAT performance (archival and in-session).*

	Archive SAT	In-Test Overall	In-Test Last 50 min	In-Test Next to last 50 min
<u>Short Condition</u>				
Pretest	.01	-.12	-.13	
Interim	.01	-.16*	-.16*	
Posttest	-.01	-.14	-.16*	
Composite Fatigue	*.00*	*.15**	*-.17**	
<u>Standard Condition</u>				
Pretest	-.03	-.11	-.05	
Interim 1	-.08	-.18**	-.15*	
Interim 2	-.03	-.13	-.12	
Posttest	-.01	-.10	-.10	
Composite Fatigue	-.05	-.15*	-.13	
<u>Long Condition</u>				
Pretest	-.03	-.14*	-.12	-.09
Interim 1	-.10	-.29**	-.29**	-.24*
Interim 2	-.08	-.23**	-.22**	-.20**
Interim 3	-.07	-.17**	-.18**	-.16**
Posttest	-.02	-.12	-.12	-.13
Composite Fatigue	*-.07*	*-.22**	*-.22**	*-.20***

Table 7. *Trait complexes and their intercorrelations*

Factor/Trait Complex

I. **Need for Achievement/Mastery**
Scales: Conscientiousness, Self-Discipline, Time and Study Environmental Management, Need for Achievement, Cautiousness, Effort Regulation, Mastery, Organization, Rehearsal
Number of scales = 9, α = .86

II. **Desire to Learn/TIE**
Scales Intrinsic Goals, Typical Intellectual Engagement, Desire to Learn, Critical Thinking, Task Value, Openness to Experience, Elaboration, Metacognitive Self-Regulation, Investigative Interests, Self-Consciousness Focus, Artistic Interests, Realistic Interest, Conventional Interests.
Number of scales = 13, α = .85

III. **Neuroticism/Anxiety**
Scales: Neuroticism, Emotionality, Worry, Trait Anxiety, Behavioral Inhibition System, Test Anxiety
Number of scales = 6, α = .89

IV. Extroversion
Scales: Extroversion (NEO-FFI), Extroversion (IPIP), Behavioral Activation System - Fun Seeking, Peer Learning, Help Seeking, Impulse Control (Reversed),
Number of scales = 6, α = .76

V. Extrinsic goals
Scales: Extrinsic Goals, Behavioral Activation System - Reward Responsiveness, Control of Learning Beliefs
Number of scales = 3, α = .63

VI. Competitiveness/Other-oriented goals
Scales: Competitiveness, Other-oriented goals, Agreeableness (Reversed), Behavioral Activation System - Drive
Number of scales = 4, α = .70

Correlations among trait complexes

	1	2	3	4	5
1. nAch/Mastery					
2. Desire to Learn/TIE	.44**				
3. Neuroticism/Anxiety	-.10	-.12			
4. Extroversion	.16*	.23**	-.19**		
5. Extrinsic Goals	.28**	.28**	.14*	.16*	
6. Competitiveness	.10	.03	.07	.26**	.28**

Notes: $N = 239$, $df = 237$, $*p < .01$; $**p < .01$; N = Neuroticism, nAch = need for achievement; DTL = Desire to learn; TIE = Typical Intellectual Engagement.

Table 8. *Correlations between trait complexes, self-concept/self-estimates of ability and SAT test performance*

	nAch/Mastery	DTL/TIE	N/Anxiety	Extroversion	Extrinsic Goals	Compet-itiveness	Math SC Self-Estimates	Verbal SC Self-Estimates
CB Records SAT (V+ Q)	-.06	.20**	-.10	.03	-.01	.15*	.38**	.19**
Short Session SAT	-.05	.17*	-.14*	.02	.03	.08	.37**	.12
Standard Session SAT	-.02	.16*	-.12	-.04	.03	.12	.40**	.06
Long Session SAT	.02	.15*	-.15*	.02	.03	.18*	.44**	.18**
Short -- Last 50 min	.00	.16*	-.11	-.00	.05	.14*	.30**	.17*
Standard -- Last 50 min	-.04	.16*	-.06	-.05	.03	.09	.33**	.14*
Long (next to last 50 min)	.04	.19**	-.17*	.02	.00	.14*	.35**	.21**
Long - Last 50 min	.04	.22**	-.10	-.03	.03	.14*	.36**	.25**

df (range from 200 to 221). $*p < .05$; $**p < .01$. V = Verbal; Q = Quantitative, DTL = Desire to learn; TIE = typical intellectual engagement; SC = Self-concept

Table 9. *Correlations between trait complexes, self-concept/self-estimates of abilities and subjective reports during testing*

	nAch/Mastery	DTL/TIE	N/Anxiety	Extroversion	Extrinsic Goals	Compet-itiveness	Math SC Self-Estimates	Verbal SC Self-Estimates
Subjective Fatigue								
Short Pretest	-.14*	-.23**	.30**	.06	-.06	.15*	-.12	.05
Short Interim	-.16*	-.18*	.34**	.07	-.03	.12	-.11	.02
Short Posttest	-.12	-.20**	.34**	.03	-.04	.15*	-.10	.08
Standard Pretest	-.29**	-.30**	.31**	.05	-.12	.11	-.12	-.03
Standard Interim 1	-.22**	-.22**	.35**	-.00	-.02	.06	-.11	-.07
Standard Interim 2	-.25**	-.21**	.35**	.00	-.04	.15*	-.11	-.04
Standard Posttest	-.20**	-.15*	.37**	-.01	-.02	.18**	-.08	-.03
Long Pretest	-.27**	-.21**	.21**	.20**	-.09	.02	-.07	-.06
Long Interim 1	-.22**	-.21**	.29**	.09	-.01	.02	-.13	-.03
Long Interim 2	-.17**	-.20**	.28**	.10	-.04	.06	-.11	-.03
Long Interim 3	-.21**	-.17*	.30**	-.00	-.03	.04	-.09	-.04
Long Posttest	-.19**	-.11	.30**	-.00	-.01	.04	-.07	-.03
Negative Affect								
Short Pretest	-.16*	-.15*	.28**	-.04	.01	.01	-.14*	-.08
Short Interim	-.20**	-.06	.30**	.00	.00	.05	-.12	-.05
Short Posttest	-.09	-.04	.28**	.04	.03	.11	-.07	-.01
Standard Pretest	-.14*	-.12	.27**	.04	-.05	.13	-.11	-.12
Standard Interim 1	-.17*	-.10	.27**	.08	.03	.10	-.05	-.10
Standard Interim 2	-.16*	-.13	.30**	.03	-.00	.10	-.10	-.11
Standard Posttest	-.15*	-.11	.27**	.01	.01	.21**	-.05	-.10
Long Pretest	-.15*	-.06	.29**	.15*	.05	.15*	-.00	-.08
Long Interim 1	-.20**	-.04	.25**	.07	.11	.08	-.02	-.02
Long Interim 2	-.13	-.06	.20**	.11	.05	.11	-.02	.00
Long Interim 3	-.07	-.02	.21**	.05	.09	.12	.00	.01
Long Posttest	-.01	.04	.24**	.10	.05	.16*	.03	.01

Table 9 continued. *Correlations between trait complexes, self-concept/self-estimates of abilities and subjective reports during testing*

	nAch/Mastery	DTL/TIE	N/Anxiety	Extroversion	Extrinsic Goals	Compet-itiveness	Math SC Self-Estimates	Verbal SC Self-Estimates
Positive Affect								
Short Pretest	.11	.24**	-.14*	.03	.10	.05	.18**	-.08
Short Interim	.12	.19**	-.21**	.05	-.00	-.10	.14*	-.02
Short Posttest	.09	.25**	-.17*	.15*	.07	-.03	.06	-.06
Standard Pretest	.06	.23**	-.16*	.11	.09	.05	.19**	-.09
Standard Interim 1	.07	.24**	-.24**	.08	.06	.00	.16*	-.00
Standard Interim 2	.18**	.29**	-.26**	.12	.10	.00	.20**	-.00
Standard Posttest	.18**	.31**	-.18**	.20**	.14*	-.02	.13	.01
Long Pretest	.21**	.22**	-.06	.04	.15*	.07	.13*	-.03
Long Interim 1	.19**	.20**	-.14*	.09	.06	.01	.19**	-.04
Long Interim 2	.18**	.21**	-.13	.09	.05	.07	.20**	-.05
Long Interim 3	.24**	.28**	-.15*	.22**	.11	.10	.19**	.01
Long Posttest	.14*	.19**	-.14*	.20**	.04	.07	.12	.05
Positive Motivation								
Short Pretest	.24**	.15*	.05	-.09	.15*	-.07	.04	.06
Short Interim	.20**	.12	-.00	-.06	.04	-.10	.06	.02
Short Posttest	.18**	.10	-.02	.01	.08	-.07	.15*	.03
Standard Pretest	.16*	.09	-.01	-.04	.06	-.05	.02	.07
Standard Interim 1	.18**	.19**	-.05	.00	.12	-.01	.09	.09
Standard Interim 2	.24**	.20**	-.08	.00	.12	-.00	.09	.10
Standard Posttest	.20**	.14*	-.10	-.02	.09	-.05	.09	.08
Long Pretest	.33**	.13	.00	-.02	.06	-.03	.00	.08
Long Interim 1	.34**	.16*	-.06	.03	.05	-.03	.06	.09
Long Interim 2	.24**	.12	.01	-.01	.07	-.08	.03	.01
Long Interim 3	.25**	.18**	-.04	.04	.07	-.10	.10	.04
Long Posttest	.29**	.20**	-.04	.07	.11	-.04	.13	.13

Table 9 continued. *Correlations between trait complexes, self-concept/self-estimates of abilities and subjective reports during testing*

	nAch/Mastery	DTL/TIE	N/Anxiety	Extroversion	Extrinsic Goals	Compet-itiveness	Math SC Self-Estimates	Verbal SC Self-Estimates
Confidence								
Short Pretest	.19**	.29**	-.33**	.04	.12	.00	.31**	.14*
Short Interim	.21**	.24**	-.35**	.00	.03	-.02	.31**	.15*
Short Posttest	.20**	.25**	-.35**	.10	.04	.00	.29**	.17*
Standard Pretest	.22**	.24**	-.36**	.04	.12	.04	.30**	.18**
Standard Interim 1	.24**	.28**	-.38**	.07	.09	.05	.34**	.18**
Standard Interim 2	.26**	.29**	-.37**	.07	.15*	.02	.32**	.18**
Standard Posttest	.23**	.26**	-.37**	.10	.14*	.04	.33**	.17*
Long Pretest	.28**	.34**	-.31**	.02	.19**	.09	.33**	.23**
Long Interim 1	.28**	.26**	-.31**	.08	.03	.10	.28**	.21**
Long Interim 2	.24**	.28**	-.31**	.05	.04	.10	.31**	.20**
Long Interim 3	.23**	.28**	-.32**	.13	.11	.08	.30**	.21**
Long Posttest	.25**	.23**	-.33**	.11	.10	.11	.32**	.25**
Self-Efficacy								
Short	.05	.22**	-.32**	.10	.04	.18*	.46**	.28**
Standard	.05	.19**	-.30**	.06	.03	.24**	.46**	.26**
Long	.08	.20**	-.32**	.08	.04	.21**	.48**	.34**

Table 10. *Summary of hierarchical regressions for predicting Subjective Fatigue during testing (with SAT performance at last step)*

Model 1: Trait Complexes, Self-Concept and Self-estimates of abilities in Step 1. Pretest Subjective Fatigue in Step 2.

		Step 1 Trait complexes & SC/SEAB	Step 2 Pretest Fatigue	Final Step SAT performance
Short Pretest	R^2 to add	.22**		
	Total R^2	.22**		
Short Posttest	R^2 to add	.25**	.30**	.00
	Total R^2	.25**	.55**	.55**
Standard Pretest	R^2 to add	.29**		
	Total R^2	.29**		
Standard Posttest	R^2 to add	.24**	.29**	.00
	Total R^2	.24**	.53**	.53**
Long Pretest	R^2 to add	.30**		
	Total R^2	.30**		
Long Posttest	R^2 to add	.22**	.14**	.01
	Total R^2	.22**	.36**	.37**

Model 2: Step 1: Trait Complexes, Self-Concept and Self-estimates of abilities. Step 2: Posttest Subjective Fatigue in Short Condition; Step 3: Pretest Subjective Fatigue.

		Step 1 Trait complexes & SC/SEAB	Step 2 Short Cond Posttest Fatigue	Step 3 In-session Pretest Fatigue	Final Step SAT performance
Standard Pretest	R^2 to add	.29**	.15**		.00
	Total R^2	.29**	.44**		.44**
Standard Posttest	R^2 to add	.24**	.33**	.09**	.00
	Total R^2	.24**	.57**	.66**	.66**
Long Pretest	R^2 to add	.30**	.11**		.00
	Total R^2	.30**	.41**		.41**
Long Posttest	R^2 to add	.22**	.24**	.04**	.01
	Total R^2	.22**	.46**	.50**	.51**

**$p < .01$. Step 1 df = 8,151; Step 2 df = 1,150; Step 3 df = 1, 149

Note: SC/SEAB = Self-concept and Self-estimates of abilities

Table 11. *Summary of hierarchical regressions for predicting SAT performance for short, standard, long conditions from trait complex scores, self-concept/self-estimates of ability, and subjective fatigue during testing.*

		Step 1 Trait complexes	Step 2 SC/SEAB	Final Step Subjective Fatigue
Short	R^2 to add	.10**	.13**	.02ns
	Total R^2	.10**	.23**	.25**
Standard	R^2 to add	.09*	.14**	.02ns
	Total R^2	.09*	.23**	.24**
Long (last 50min)	R^2 to add	.12**	.16**	**.04***
	Total R^2	.12**	.28**	.32**
Long (next to last 50min)	R^2 to add	.11**	.14**	**.06***
	Total R^2	.11**	.25**	.31**

Step 1 df = 6, 155 (short); 6, 155 (standard); 6, 158 (long)
Step 2 df = 2, 153 (short); 2, 153 (standard); 2, 156 (long)
Step 3 df = 2, 151 (short), 3, 150 (standard); 4, 152 (long)

Table 12. *Correlations among subjective measure composites*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Subjective Fatigue - Short														
2. Subjective Fatigue - Standard	.68**													
3. Subjective Fatigue - Long	.61**	.70**												
4. Negative Affect - Short	.54**	.43**	.35**											
5. Negative Affect - Standard	.45**	.60**	.44**	.72**										
6. Negative Affect - Long	.29**	.43**	.50**	.56**	.67**									
7. Positive Affect - Short	-.43**	-.25**	-.25**	.01	.01	.02								
8. Positive Affect - Standard	-.33**	-.43**	-.25**	-.00	.01	.04	.72**							
9. Positive Affect - Long	-.25**	-.29**	-.37**	.02	.03	-.01	.72**	.80**						
10. Positive Motivation - Short	-.28**	-.23**	-.16*	-.18**	-.12	-.11	.38**	.30**	.27**					
11. Positive Motivation - Standard	-.21**	-.25**	-.20**	-.11	-.09	-.13	.29**	.37**	.28**	.81**				
12. Positive Motivation - Long	-.19*	-.27**	-.27**	-.08	-.14	-.19**	.33**	.37**	.35**	.80**	.87**			
13. Confidence - Short	-.66**	-.54**	-.46**	-.50**	-.39**	-.34**	.32**	.31**	.22**	.40**	.31**	.29**		
14. Confidence - Standard	-.48**	-.74**	-.56**	-.35**	-.45**	-.37**	.21**	.42**	.28**	.34**	.43**	.41**	.76**	
15. Confidence - Long	-.41**	-.57**	-.71**	-.31**	-.36**	-.44**	.19**	.30**	.34**	.25**	.32**	.40**	.72**	.84**

$^{*}p < .05$; $^{**}p < .01$

Table 13. *Correlations between composite subjective measures and trait complexes and math and verbal self-concept and self-estimates of ability.*

	Trait Complex						Math	Verbal
	1	2	3	4	5	6	SC/SEAB	SC/SEAB
Subjective Fatigue								
Short	-.13	-.18**	.35**	.06	-.02	.15*	-.10	.06
Standard	-.23**	-.20**	.38**	-.00	-.03	.12	-.11	-.06
Long	-.22**	-.20**	.33**	.04	-.03	.04	-.11	-.05
Negative Affect								
Short	-.15*	-.05	.30**	.02	.02	.08	-.10	-.03
Standard	-.17*	-.12	.30**	.04	.01	.15*	-.07	-.11
Long	-.10	-.05	.26**	.08	.06	.11	-.01	-.01
Positive Affect								
Short	.12	.26**	-.19**	.09	.06	-.02	.14*	-.06
Standard	.14*	.30**	-.24**	.15*	.12	.01	.19**	-.01
Long	.19**	.28**	-.16*	.21**	.10	.10	.20**	-.05
Positive Motivation								
Short	.21**	.12	.01	-.05	.08	-.09	.08	.03
Standard	.21**	.17*	-.07	-.01	.10	-.03	.08	.09
Long	.27**	.16*	-.00	.00	.08	-.09	.06	.05
Confidence								
Short	.21**	.27**	-.38**	.04	.06	-.01	.32**	.16*
Standard	.27**	.30**	-.41**	.07	.13	.03	.35**	.20**
Long	.27**	.29**	-.36**	.09	.08	.08	.34**	.23**

Table 14. *Correlations between composite self-report measures and SAT performance.*

		Overall			Last 50 min			next to last
	Archive	Short	Standard	Long	Short	Standard	Long	50 min Long
Subjective Fatigue - Short	-.00	-.15*	-.12	-.03	-.16*	-.09	-.07	-.11
Subjective Fatigue - Standard	-.05	-.12	-.15*	-.11	-.16*	-.13	-.11	-.13
Subjective Fatigue - Long	-.07	-.19**	-.16*	-.21**	-.20**	-.14*	-.21**	-.19**
Negative Affect - Short	-.13	-.28**	-.19**	-.18*	-.31**	-.19**	-.15*	-.25**
Negative Affect - Standard	-.17*	-.25**	-.22**	-.20**	-.29**	-.23**	-.20**	-.27**
Negative Affect - Long	-.15*	-.21**	-.19**	-.17*	-.22**	-.16*	-.16*	-.20**
Positive Affect - Short	.01	.07	.02	.00	.00	-.04	.01	.03
Positive Affect - Standard	.09	.07	.11	.08	.03	.06	.07	.05
Positive Affect - Long	.16*	.14	.14	.14*	.12	.08	.16*	.14*
Positive Motivation - Short	-.02	.12	.08	.06	.14*	.11	.02	.06
Positive Motivation - Standard	.01	.12	.13	.11	.15*	.13	.08	.07
Positive Motivation - Long	.06	.15*	.12	.16*	.16*	.13	.14*	.13
Confidence - Short	.28**	.41**	.32**	.32**	.37**	.28**	.32**	.32**
Confidence - Standard	.27**	.34**	.35**	.36**	.31**	.30**	.31**	.31**
Confidence - Long	.37**	.43**	.42**	.50**	.41**	.39**	.47**	.43**

Table 15. *Correlations among composite self-report measures (across Short, Standard, and Long conditions).*

	1	2	3	4
1. Fatigue				
2. Negative Affect	.58**			
3. Positive Affect	-.41**	.03		
4. Positive Motivation	-.28**	-.15*	.40**	
5. Confidence	-.70**	-.50**	.36**	.40**

df = 175

Table 16. *Summary of hierarchical regressions for predicting composite (across Short, Standard, and Long conditions) Subjective Fatigue during testing (with SAT performance at last step)*

		Step 1 Trait complexes	Step 2 SC/SEAB	Step 3 SAT performance
Subjective Fatigue	R^2 to add	.31**	.01	.00
	Total R^2	.31**	.32**	.32**
Negative Affect	R^2 to add	.18**	.00	.04**
	Total R^2	.18**	.18**	.22**
Positive Affect	R^2 to add	.17**	.04*	.00
	Total R^2	.17**	.21**	.21**
Positive Motivation	R^2 to add	.06	.01**	.00
	Total R^2	.06ns	.07ns	.07ns
Confidence	R^2 to add	.36**	.07**	.02*
	Total R^2	.36**	.43**	.45**

Step 1 df = 6, 150
Step 2 df = 2, 148
Step 3 df = 1, 147

Figure Captions

Figure 1. SAT Test performance (number correct minus a fraction for number wrong) by test-length condition for the final 50 min section (Verbal [20 min] Quantitative [20 min] + Writing [10 min] test components). For the Long test length condition, both the last 50 min section and a parallel penultimate 50 min section are shown. Error bars indicate 95% confidence intervals for the means.
Figure 2. Mean subjective fatigue scale scores for pretest, interim, and posttest administrations, by test condition.
Figure 3. Mean negative affect scale scores for pretest, interim, and posttest administrations, by test condition.
Figure 4. Mean positive affect scale scores for pretest, interim, and posttest administrations, by test condition.
Figure 5. Mean positive motivation scale scores for pretest, interim, and posttest administrations, by test condition.
Figure 6. Mean confidence scale scores for pretest, interim, and posttest administrations, by test condition.
Figure 7. Composite subjective fatigue, as a function of response to effort strategy question and test condition.
Figure 8. SAT test performance during the final 50 min module, as a function of response to effort strategy question and test condition.
Figure 9. Mean response to the question of “It was hard to keep up my motivation on the SAT” as a function of response to effort strategy question and test condition.

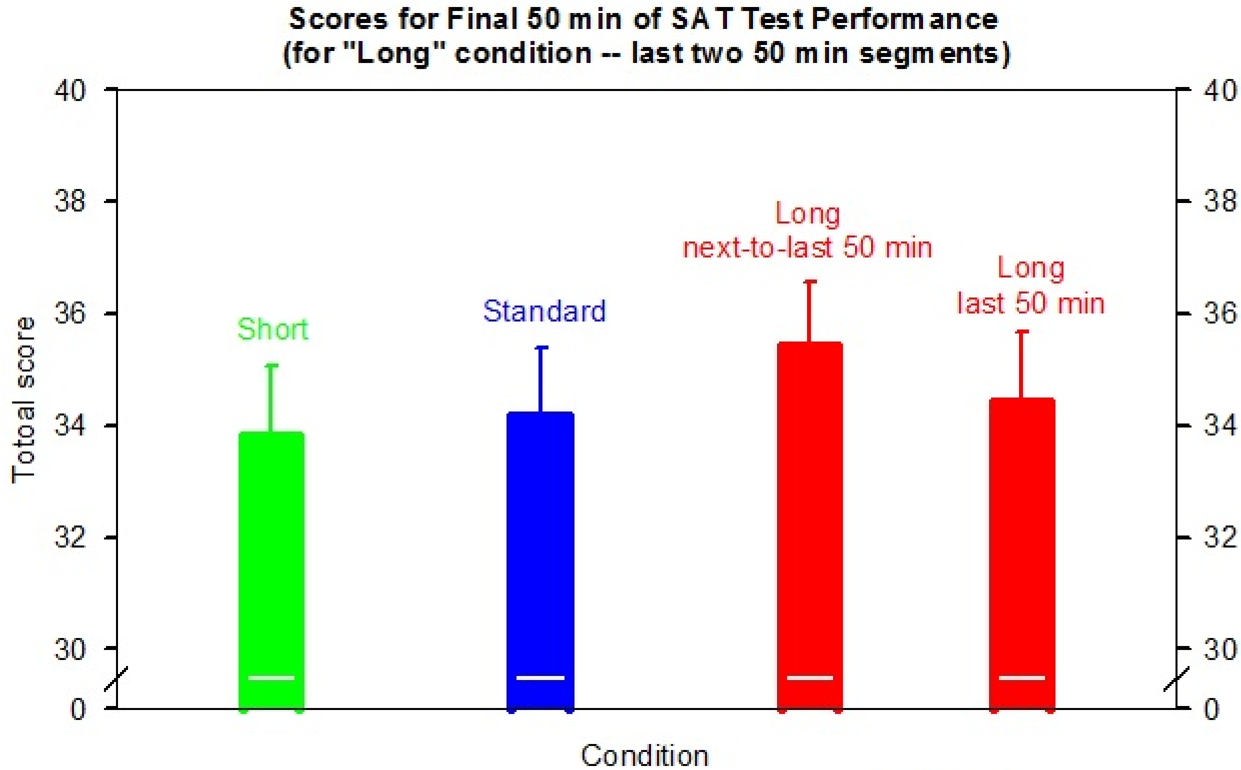

Scores for Final 50 min of SAT Test Performance
(for "Long" condition -- last two 50 min segments)
40
38
36
34
32
30
0
Totoal score
Short
Standard
Long
next-to-last 50 min
Long
last 50 min
Condition

Figure 1

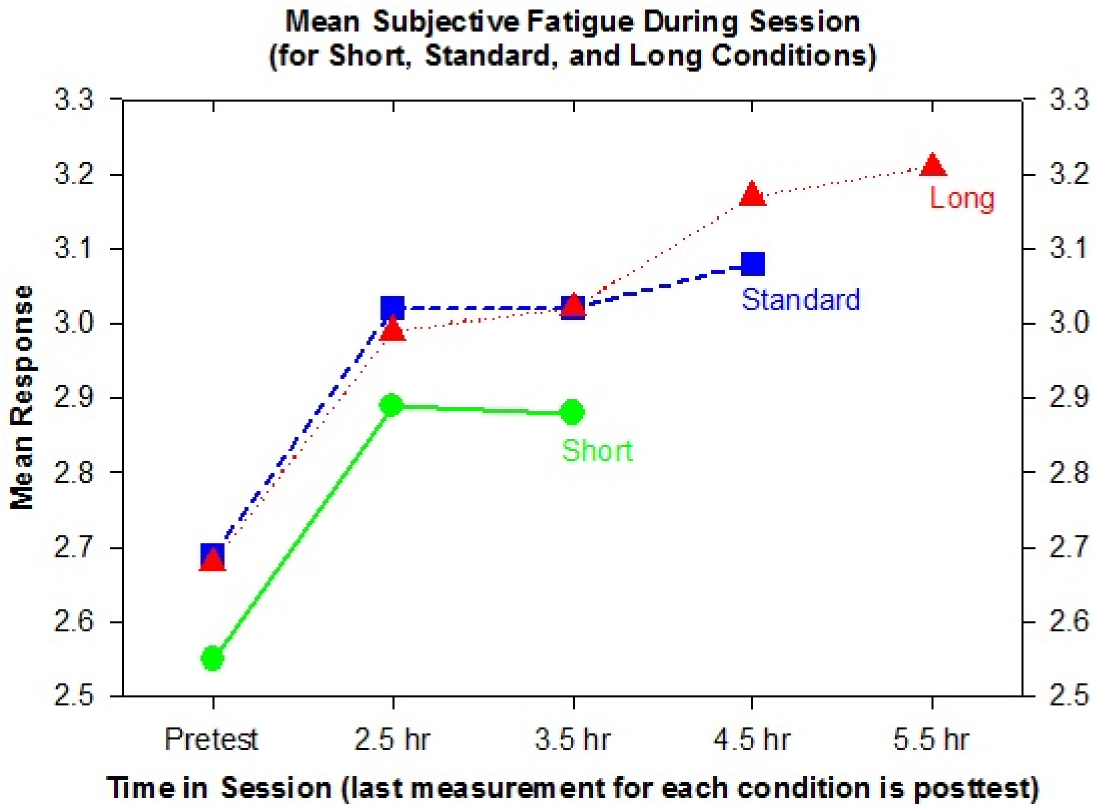

Mean Subjective Fatigue During Session
(for Short, Standard, and Long Conditions)
Mean Response
3.3
3.2
3.1
3.0
2.9
2.8
2.7
2.6
2.5
Long
Standard
Short
Pretest
2.5 hr
3.5 hr
4.5 hr
5.5 hr
Time in Session (last measurement for each condition is posttest)

Figure 2

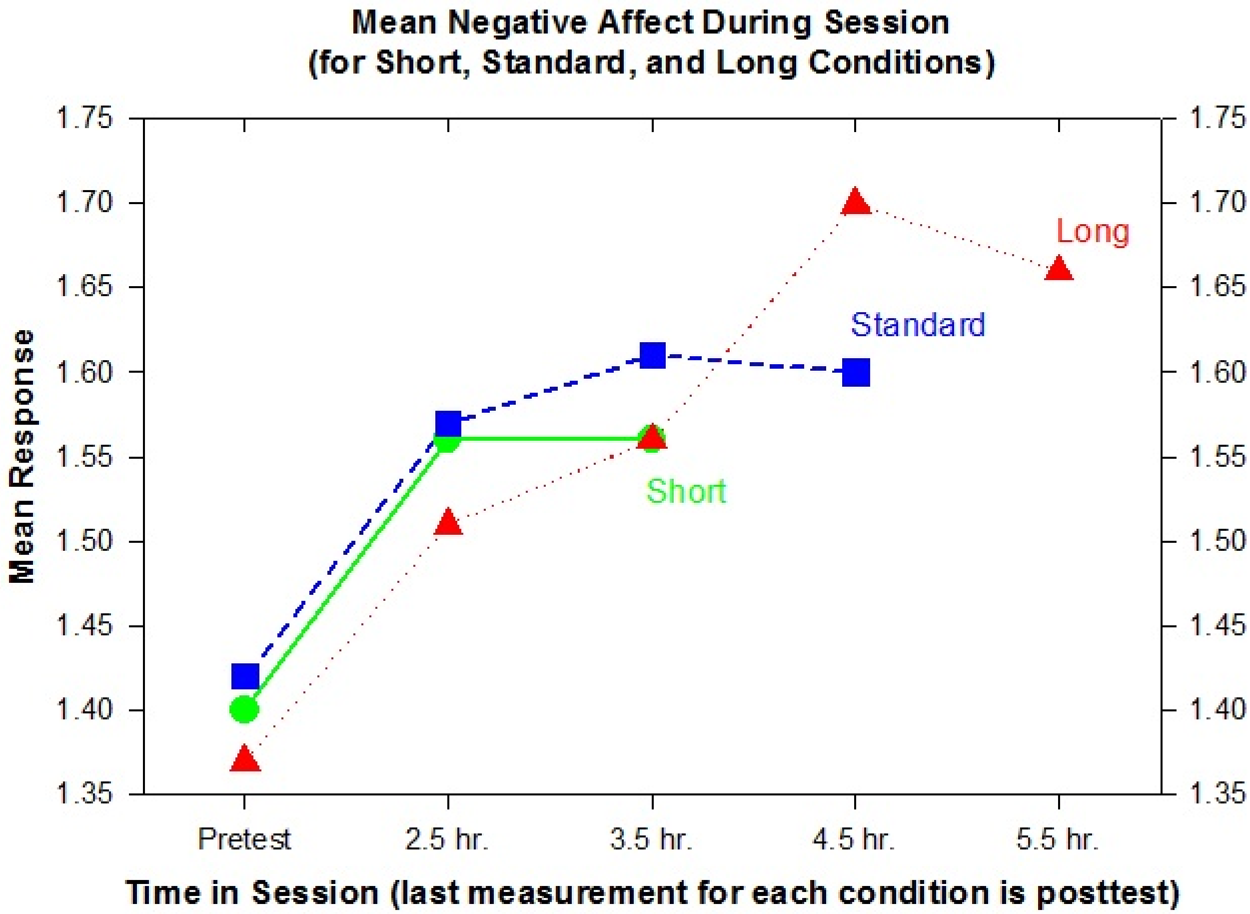

Mean Negative Affect During Session
(for Short, Standard, and Long Conditions)
Mean Response
1.75
1.70
1.65
1.60
1.55
1.50
1.45
1.40
1.35
Long
Standard
Short
Pretest
2.5 hr.
3.5 hr.
4.5 hr.
5.5 hr.
Time in Session (last measurement for each condition is posttest)

Figure 3

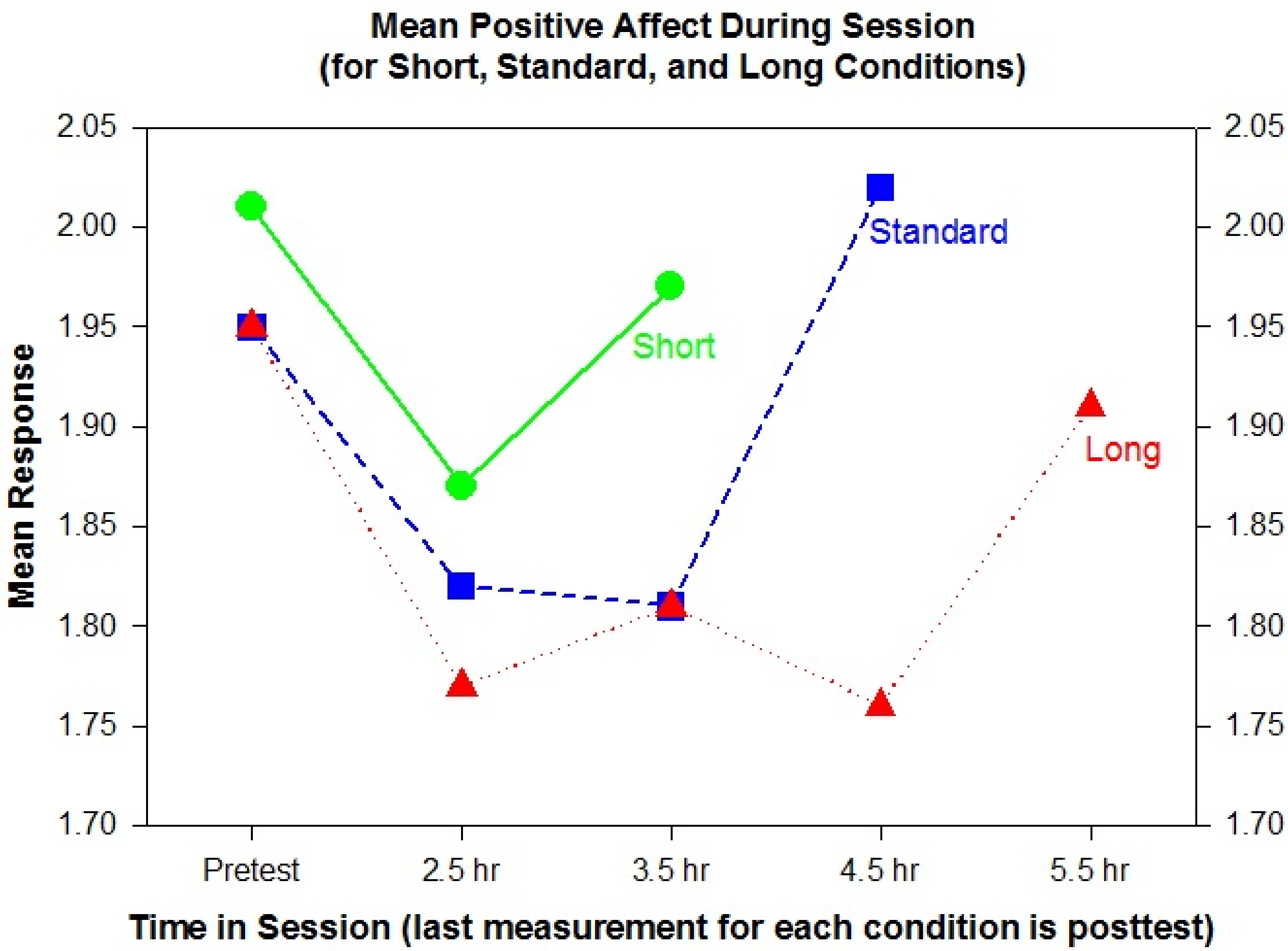

Mean Positive Affect During Session
(for Short, Standard, and Long Conditions)
2.05
2.00
1.95
1.90
1.85
1.80
1.75
1.70
Mean Response
Short
Standard
Long
Pretest
2.5 hr
3.5 hr
4.5 hr
5.5 hr
Time in Session (last measurement for each condition is posttest)

Figure 4

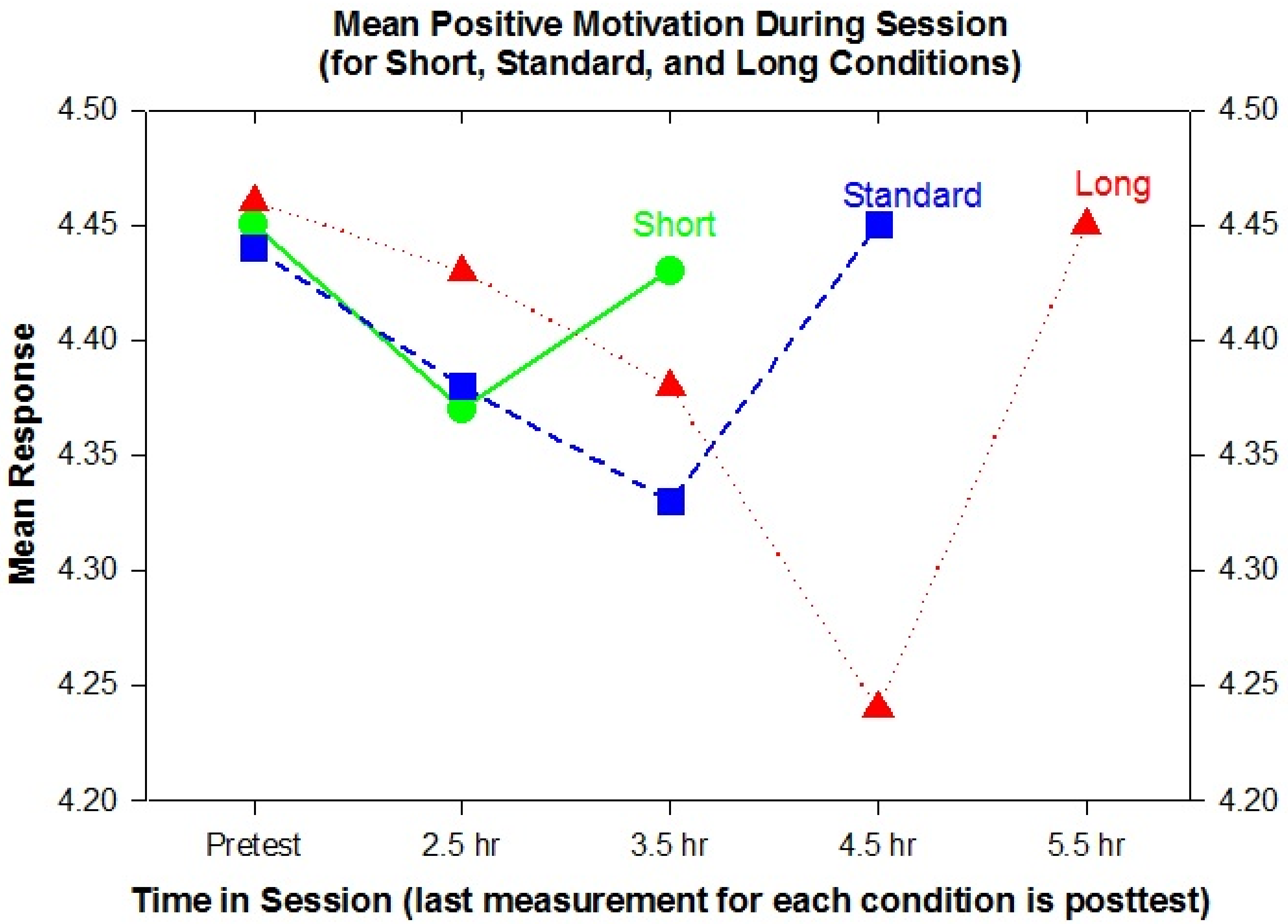

Mean Positive Motivation During Session
(for Short, Standard, and Long Conditions)
Mean Response
Time in Session (last measurement for each condition is posttest)
Pretest
2.5 hr
3.5 hr
4.5 hr
5.5 hr
4.20
4.25
4.30
4.35
4.40
4.45
4.50
Short
Standard
Long

Figure 5

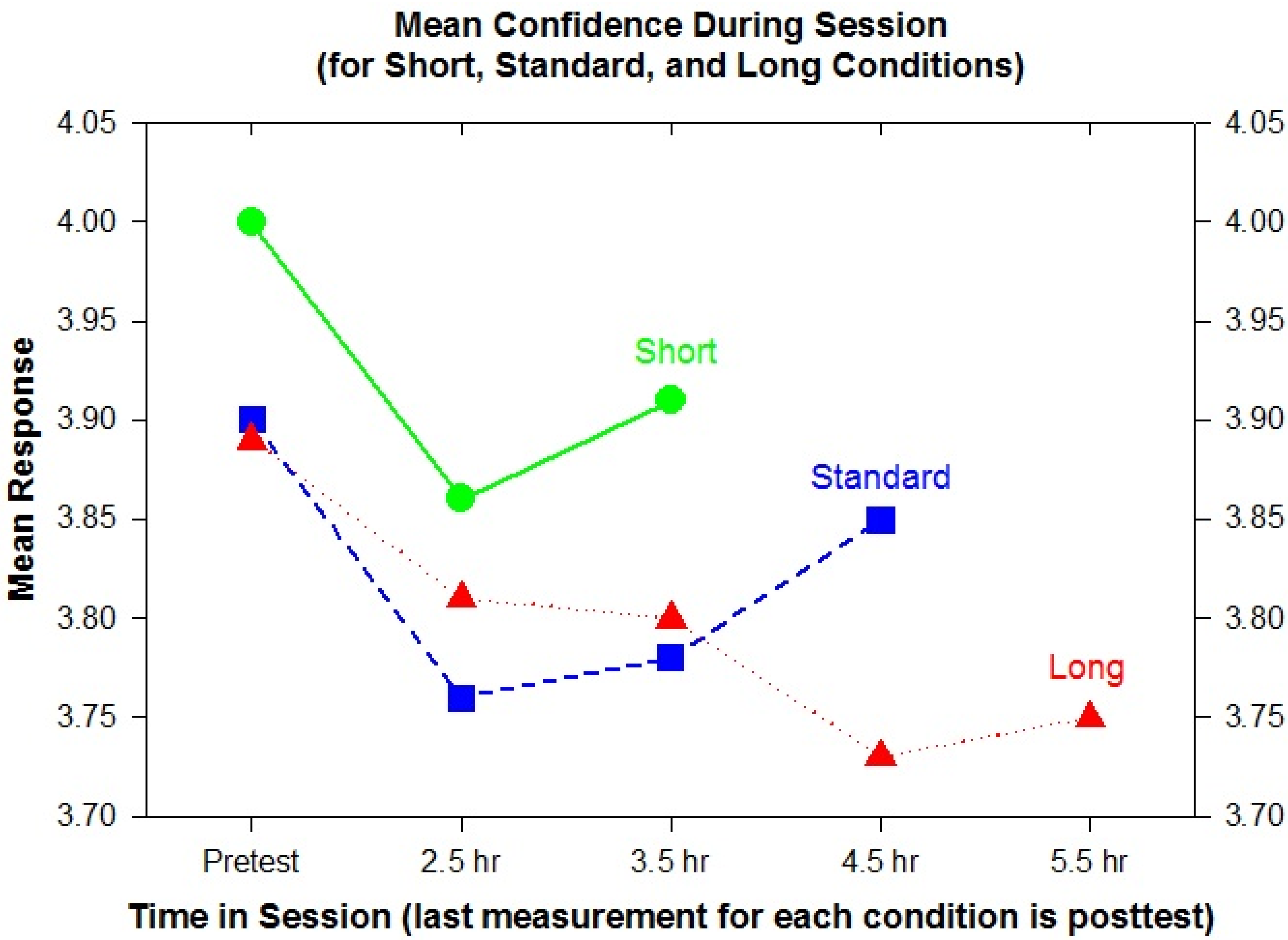
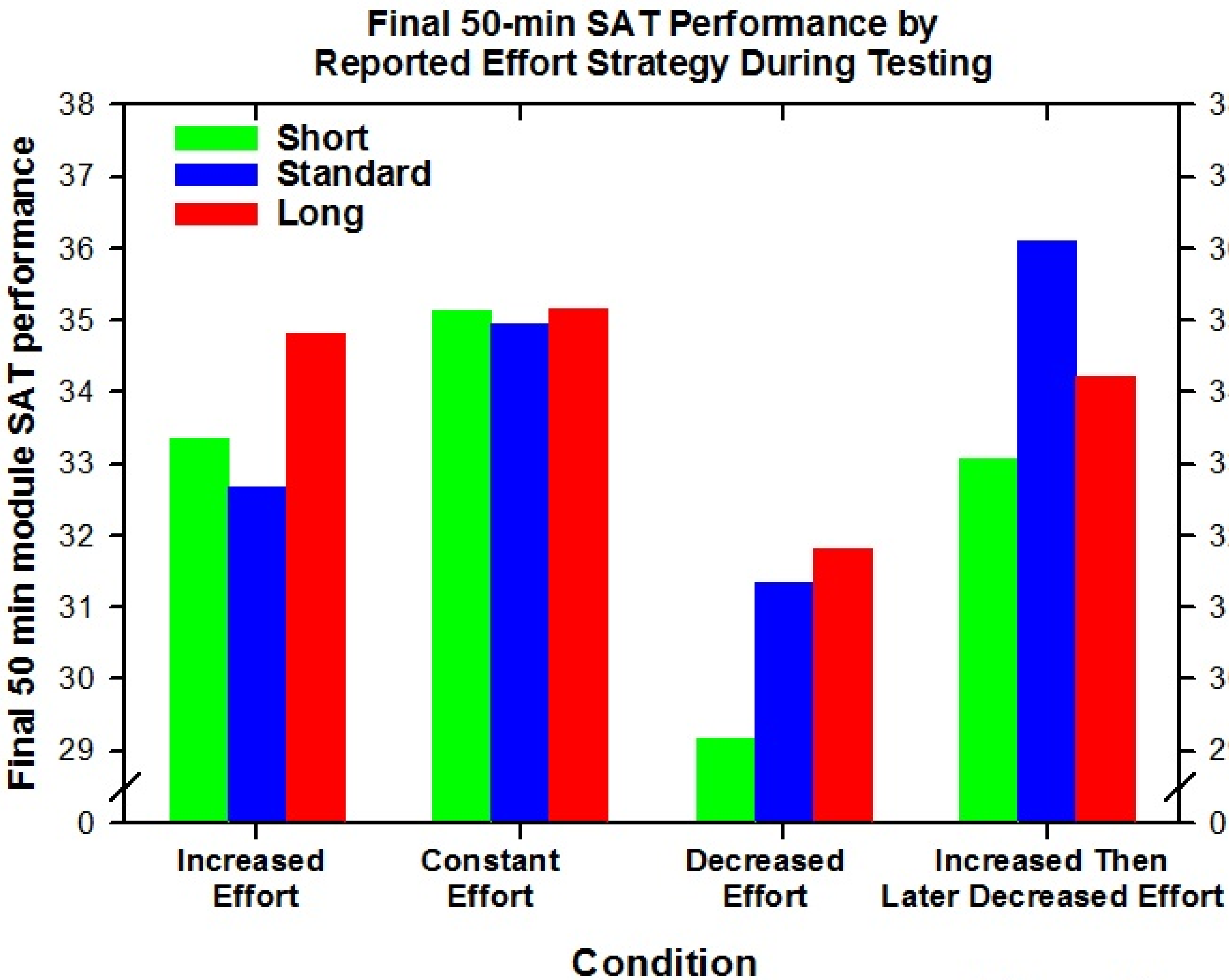

Mean Confidence During Session
(for Short, Standard, and Long Conditions)
Mean Response
4.05
4.00
3.95
3.90
3.85
3.80
3.75
3.70
Short
Standard
Long
Pretest
2.5 hr
3.5 hr
4.5 hr
5.5 hr
Time in Session (last measurement for each condition is posttest)

Figure 6



Final 50-min SAT Performance by
Reported Effort Strategy During Testing
Short
Standard
Long
Final 50 min module SAT performance
38
37
36
35
34
33
32
31
30
29
0
Increased
Effort
Constant
Effort
Decreased
Effort
Increased Then
Later Decreased Effort
Condition

Figure 7

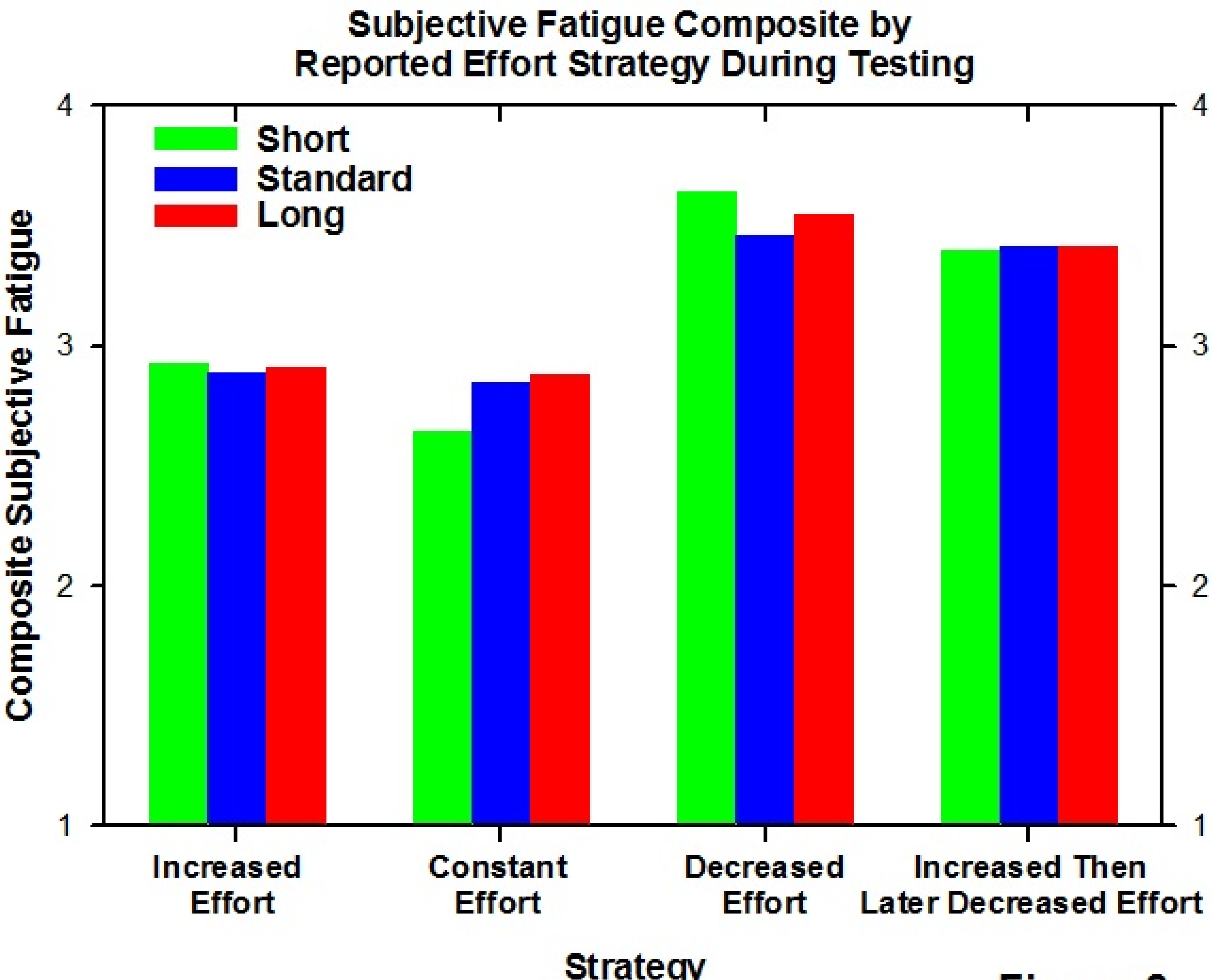
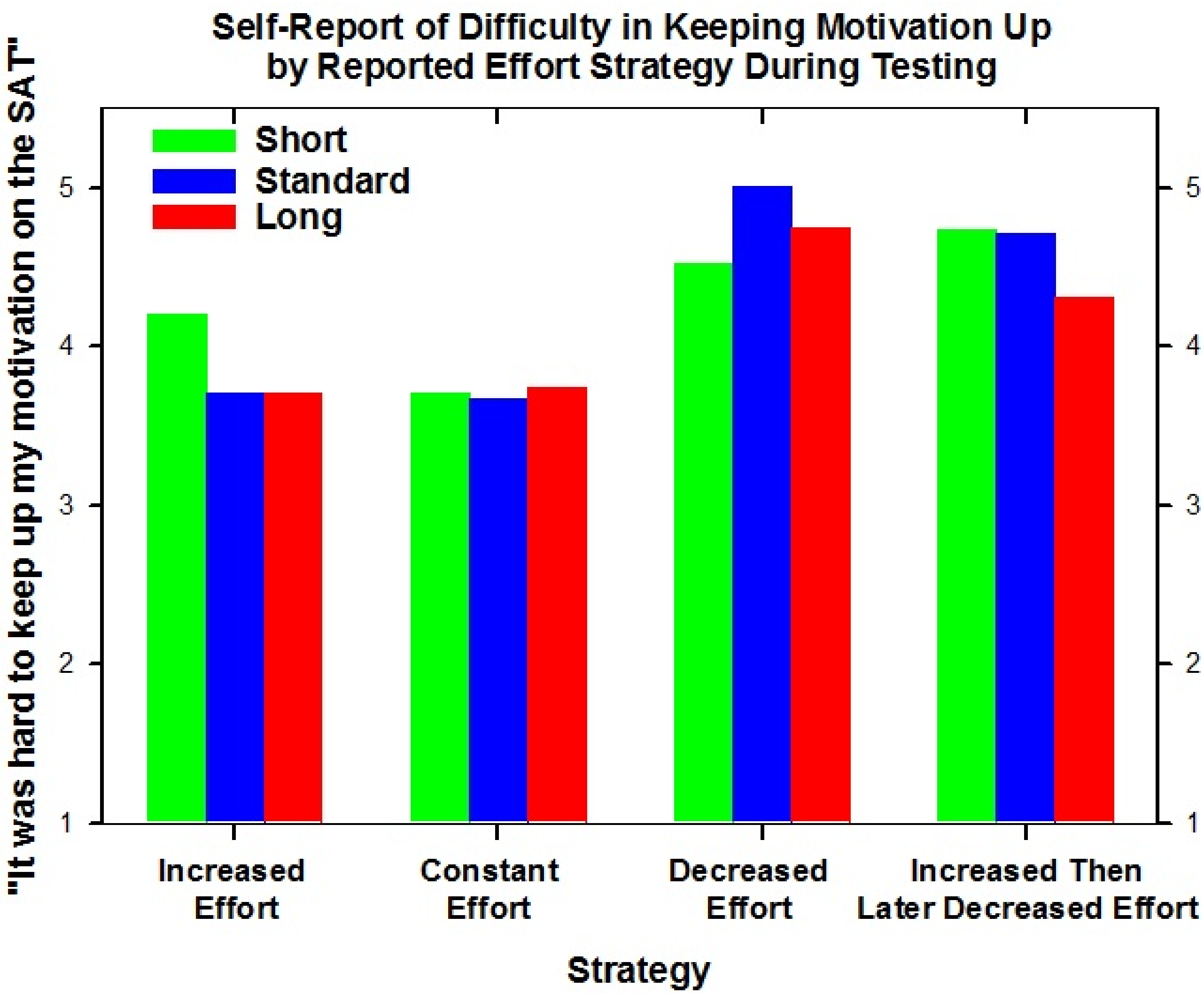

Subjective Fatigue Composite by
Reported Effort Strategy During Testing
Short
Standard
Long
Composite Subjective Fatigue
1
2
3
4
Increased Effort
Constant Effort
Decreased Effort
Increased Then Later Decreased Effort
Strategy

Figure 8



Self-Report of Difficulty in Keeping Motivation Up
by Reported Effort Strategy During Testing
"It was hard to keep up my motivation on the SAT"
Short
Standard
Long
1
2
3
4
5
Increased Effort
Constant Effort
Decreased Effort
Increased Then Later Decreased Effort
Strategy

Figure 9