Journal of Experimental Psychology: Applied
2009, Vol. 15, No. 2, 163–181

1076-898X/09/$12.00 DOI: 10.1037/a0015719

Test Length and Cognitive Fatigue: An Empirical Examination of Effects on Performance and Test-Taker Reactions

Phillip L. Ackerman and Ruth Kanfer
Georgia Institute of Technology

Person and situational determinants of cognitive ability test performance and subjective reactions were examined in the context of tests with different time-on-task requirements. Two hundred thirty-nine first-year university students participated in a within-participant experiment, with completely counterbalanced treatment conditions and test forms. Participants completed three test sessions of different length: (a) a standard-length SAT test battery (total time 4½ hr), (b) a shorter SAT test battery (total time 3½ hr), and (c) a longer SAT test battery (total time 5½ hr). Consistent with expectations, subjective fatigue increased with increasing time-on-task. However, mean performance *increased* in the longer test length conditions, compared with the shorter test length condition. Individual differences in personality/interest/motivation trait complexes were found to have *greater* power than the test-length situations for predicting subjective cognitive fatigue before, during, and at the end of each test session. The relative contributions of traits and time-on-task for cognitive fatigue are discussed, along with implications for research and practice.

Keywords: fatigue, ability, personality, trait complex

What is cognitive fatigue? Researchers have taken several different approaches to the definition of mental/cognitive fatigue over the past 100 or so years. Some early researchers (e.g., Ebbinghaus, 1896–1897) were mainly interested in the effects of prolonged work on performance of tasks that involved cognitive functions of memory, judgment, reasoning, and other typical components of intellectual abilities. However, other researchers (e.g., Muscio, 1921a, 1921b) were concerned with the subjective aspects of mental fatigue, which as we will discuss later, involve a panoply of different motivational and attitudinal factors. Some researchers (e.g., Bartley & Chute, 1947) limited use of the term fatigue to subjective fatigue, whereas others (e.g., Dodge, 1917) termed performance effects associated with fatiguing conditions as "work decrement." In the discussion that follows, we will refer to performance effects as "cognitive fatigue" and distinguish performance effects from subjective reports, which will be referred to as "subjective cognitive fatigue." Although much of the extant experimental literature in the past century has used the term "mental fatigue," and much of the clinical literature in the past three decades has used the term "cognitive fatigue," as distinguished from physical or muscle fatigue (e.g., see Mosso, 1906), we believe that the term "cognitive" is more precise in the current context of cognitive psychology than is the broader use of "mental" to refer to cognitive processes.

Cognitive fatigue is a topic of continued importance in applied psychology. In education, training, job search, career development, and job promotion contexts, individuals must succeed on high-stakes tests (e.g., SAT, Bar Exam, Medical Board Certification) or perform critical cognitively demanding tasks that take long periods of time to complete; sometimes with minimal or no breaks over a several-hour session. To perform well under such circumstances individuals must avoid off-task thoughts that potentially distract attention away from the task, persist even when the test or task holds little in the way of intrinsic interest, and continue to exert effort even when feeling cognitively fatigued.

Increasing our knowledge about cognitive fatigue has both theoretical and practical implications. Although there is a substantial literature on trait determinants of behavior and attitudes, there is relatively little research that addresses how stable individual differences in personality and other traits relate to subjective reactions under cognitively fatiguing conditions. In addition, there is no research that addresses whether the relations between traits and subjective cognitive fatigue are stable or change with increasing time-on-task. Most notable is the lack of evidence that compares situation and trait determinants of cognitive fatigue within the same experimental framework. Delineation of the influences on cognitive fatigue has direct implications for the development and structure of high-stakes standardized tests. Such implications might relate to the consequences of time-on-task changes and testing-condition modifications in high-stakes performance contexts, and prediction of who might benefit most from such modifications. From a practical perspective, such research might also provide information

---

Phillip L. Ackerman and Ruth Kanfer, School of Psychology, Georgia Institute of Technology.

Neil Charness served as action editor on this article. A brief presentation of the results of this study is reported in College Board Research Notes #RN-37. 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. We are particularly grateful for the support provided by Wayne Camara, Krista Mattern, and Glenn Milewski. In addition, we wish to acknowledge the assistance of Stacey Wolman, Sunni Haag, and other members of the Knowledge and Skill Lab staff at Georgia Tech in recruitment, administering the tests, coding data, and other critical activities. The ideas expressed in this article are those of the authors and do not reflect the opinions or position of the College Board.

Correspondence concerning this article should be addressed to Phillip L. Ackerman, School of Psychology, Georgia Institute of Technology, 654 Cherry Street, Atlanta, GA 30332-0170. E-mail: phillip.ackerman@psych.gatech.edu

for the design of work that involves sustained cognitive effort over extended time periods. Using a person-situation interactionist framework, we investigate the independent and joint influence of person (i.e., traits) and situation (i.e., time-on-task) influences on perceptions of cognitive fatigue during ability testing.

To date, most experimental research on the effects of time-on-task and cognitive fatigue has occurred in the laboratory or the classroom, with either repetitive reaction time or mathematical tasks, or has involved the special case of vigilance tasks (e.g., monitoring displays for low frequency targets in high frequency noise; see See, Howe, Warm, & Dember, 1995, for a review). Fewer studies have been conducted in the context of high-stakes cognitive aptitude/ability testing; that is, when the results of testing have direct and powerful ramifications for an individual's educational opportunities (in the case of the SAT, ACT, or GRE tests) or access to a profession or career path (in the case of professional certification tests). The pervasive use of tests in the education and work worlds is a critical reason for investigations of cognitive fatigue in the context of high-stakes testing. For example, the current SAT test, completed by nearly 1.5 million test-takers in 2007 (College Board, 2007), involves 3 hr and 45 min of testing over a 4½-hr session.

Many professional certification tests involve even longer test times. Among people seeking to pursue a law career, admission to the Bar (the professional certification process for lawyers in the United States) requires an examination that typically involves 2 to 3 consecutive days of testing, with 6 hr/day or more of tests, depending on the state. Similarly, nearly 175,000 people annually complete one of the three, 8-hr tests conducted over a single 10-hr period, in order to achieve the status of Chartered Financial Analyst (CFA Institute, 2008; The Economist, 2008).

Surprisingly, it is unclear whether and how test length and test-taker cognitive fatigue influence performance. The few studies that have addressed test length and performance do not directly address cognitive fatigue because, for example, they have examined performance on tests with a fixed number of items, but increases in testing time (e.g., as would be found when examinees are given extra time because of special accommodations). Under such circumstances, performance inevitably increases to a greater or lesser degree, because there is a monotonically increasing probability of answering a question correctly on a speeded test, as the time allowed for solution of each item increases (e.g., see Laitusis, Morgan, Bridgeman, Zanna, & Stone, 2007).

A small sample, between-subjects procedure study by Liu, Feigenbaum, Oh, and Burton (2004) did investigate the performance and subjective fatigue effects of adding a 25-min essay section to the then-current SAT test. In general, they found no significant effects on performance in the longer test conditions, but more participants reported higher levels of subjective fatigue in the longer test conditions. The authors reported that the results from this study were compromised by the small sample size and the "nonrepresentative" nature of the sample. However, if, as popularly believed, longer tests result in poorer overall performance, such tests may yield an underestimate of the individual's capabilities, especially if similar levels of cognitive fatigue are not encountered in the criterion job/task. In addition, if some individuals are more susceptible to cognitive fatigue effects than others, and these differences are orthogonal to criterion job/task performance, then the fatiguing aspects of the test may obscure the relationship between test performance and criterion job/task performance, resulting in lower criterion-related validity. Moreover, in the case of the SAT, the recent change in test format to include an additional 45 min of testing for an essay component has resulted in numerous complaints from *test-takers* and others (e.g., FairTest, 2006; FOXNews.Com, 2005; Hildebrand, 2007; T. Lewin, 2005; MacDonald, 2005). Such concerns essentially involve one or more of following lines of reasoning: (1) Performance on the SAT is negatively affected by the additional total testing time; (2) Test-taker fatigue increases as a function of the increased total testing time; and by implication, (3) Test-taker fatigue is an influential factor in performance on the SAT.

The potential decrement in performance as a function of cognitive fatigue is one central concern, but subjective feelings of cognitive fatigue in terms of test-taker reactions to the test situation are also important. If potential test-takers believe that the test is fatiguing, and if they expect that their performance will be compromised, then they may avoid the test situation, look for ways to modify the test situation (such as test accommodations), seek to change organizational or public policy regarding the test, or even take a different educational or occupational path that does not require high-stakes testing. However, even though the test situation may give rise to feelings of subjective fatigue, individual differences in nonability traits, such as personality, interest, self-concept, and motivation may be more important determinants of individual differences in a variety of reactions to the testing situation, including subjective fatigue (e.g., see Arvey, Strickland, Drauden, & Martin, 1990; Chan, Schmitt, DeShon, Clause, & Delbridge, 1997).

Prior Research on Time-on-Task, Cognitive Fatigue, and Performance

Questions about the effects of time-on-task and cognitive fatigue on performance during testing have been raised repeatedly over the last hundred years of modern assessment research and practice. Indeed, 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 over the course of the school day on attention and performance. In a recent review of the research literature investigating time-on-task effects on cognitive fatigue and performance, Ackerman and Kanfer (2006) found that although there was a substantial older research literature on the performance effects associated with increasing the amount of time required by continuous demands on cognitive processing, the findings were inconsistent with respect to the effects of tests that last 4 hr or more.

An early study by Martyn (1913) illustrates the differences in reactions to fatiguing conditions. In that study, three participants performed mental multiplication over a 1-hr work period. One participant showed no substantial changes in task performance or other measures of fatigue as time-on-task increased. A second participant showed improved performance over the 1-hr work period, and the third participant showed lower performance, distracting thoughts, and detrimental effects on muscles, pulse, and sensory threshold. The reactions of these three participants are prototypical of behaviors and reaction patterns observed in other studies. Arai (1912), in a now-classic study, found longer solution times to mental multiplication test items as time-on-task increased to 12 hr, a finding that was later replicated by Huxtable, White, and McCartor (1946). In contrast, a

series of studies by Carmichael and his colleagues (e.g., Carmichael & Dearborn, 1947) indicated that reading performance was largely unchanged over a period of up to 6 hr of time-on-task.

Most studies investigating time-on-task effects have focused on the effects of specific situation and task characteristics on subjective cognitive fatigue and performance, rather than process explanations that accord individuals an active role in self-managing cognitively demanding performances over time. Davis (1946) provided a noteworthy exception to this trend in his study investigating time-on-task changes in performance among individuals engaged in a 70-min pilot simulation task designed to be demanding and fatiguing. Davis (1946) identified three different performance patterns, or strategies, associated with increasing time-on-task. The first pattern of performance, characteristic of nearly 75% of study participants, is represented by stable performance throughout the session. Davis noted that participants with this pattern maintained a constant level of attentional effort, regardless of the subjective experience of fatigue. The second pattern of performance observed by Davis (1946) was characterized by an increase in performance as time-on-task increased, again regardless of self-reports of cognitive fatigue. In contrast to the first two patterns, the third performance pattern noted was associated with decreasing levels of performance as time-on-task increased. Davis suggested that this pattern might be expected from individuals who perceived themselves to be fatigued, and who reduced their task-related effort to protect against further increases in cognitive fatigue; in essence, psychologically withdrawing from the task (cf., Lewin, 1935).

Davis (1946) also suggested a potential fourth pattern of performance that might occur, although it was not observed among participants in his study. Specifically, he suggested that during the early phase of time-on-task, individuals might respond to increases in perceptions of cognitive fatigue by exerting more effort and so increase performance. As time-on-task continued, however, exhaustion of effort resources might reduce effort and performance below initial baseline level. Davis' (1946) explanation of performance patterns in terms of the use of different self-regulatory strategies over time is particularly useful for understanding the lack of consistency in findings relating time-on-task to performance. Specifically, Davis' identification of different performance strategies suggests that the strategies used by participants may operate to produce constant, increasing, decreasing, or initially increasing, then decreasing mean levels of performance over time.

Perceived (Subjective) Fatigue

Although the literature does not reveal consistent patterns of performance effects related to cognitive fatigue, increases in subjective fatigue as time-on-task increases in cognitively demanding tasks are nearly ubiquitous. Short-term laboratory studies (e.g., 6 or fewer hr time-on-task) show patterns of increasing subjective fatigue as a function of increasing time-on-task. Relatively recent findings, such as that by D'Huyvetter (1987) in a study of complex information processing task performance over an extended period of time-on-task, suggest that the single most influential variable on subjective feelings of fatigue may be the total time-on-task. In D'Huyvetter's study, time-on-task was an even more important influence on subjective fatigue than whether the task involved high or low levels of total workload (i.e., effort demanded or expended).

Other studies have relied on evaluation of related constructs of subjective fatigue (such as task aversion, motivation, interest, etc.). For example, Boksem, Meijman, and Lorist (2005), in a study of 3 hr continuous time-on-task in a visual attention task, found that task aversion, measured on a scale from zero to 10, increased from a mean of 1.0 to a mean of 8.6. Similar increases in task aversion over 2 hr time-on-task in a task switching paradigm were found by Lorist et al. (2000). In addition, several studies using measures of mood indicate that feelings of "vigor" decline during the experimental session. Kaneko and Sakamoto (2001), in a study of simple addition over a 6-hr task, found increases in subjective fatigue on three different groups of adjective checklist items. Although there was a general increase in endorsement of fatigue items, the group of items associated with "Drowsiness and dullness" was endorsed with much greater frequency than the items associated with "Difficulty in concentration due to mental fatigue" and "Physical fatigue of body parts." When the task involves a high level of cognitive effort, subjective fatigue tends to correspond to changes in affect and physiological reports, including higher levels of drowsiness, difficulty in concentration, and other physical manifestations (such as aches in eyes, shoulders, and neck; e.g., see Kinsman & Weiser, 1976).

The difficulty in integrating the research on subjective fatigue on the one hand, with the research on performance effects in fatiguing conditions on the other hand, is that on many occasions there is no concomitant decline in mean performance, even in the face of substantial increases in subjective fatigue. Several researchers have attempted to explain why this pattern of results tends to occur. Prominent among these explanations is Thorndike's (1900) observation that individuals tend to feel fatigued before their performance is affected; in which case performance effects might only be expected if time-on-task proceeds far longer than is required to observe increases in subjective fatigue. Similarly, Myers (1937) suggested that subjective fatigue rather has a protective value for the individual (as an indication that he or she should discontinue the activity) and might not in fact have any apparent influence on actual performance.

Perhaps most relevant to these viewpoints is the noted dissociation between various measures of subjective workload and performance that has been found in the experimental and ergonomics literature (e.g., see Yeh & Wickens, 1988). According to these authors, perceived workload provides an index of how much effort the individual is expending, such that if maintaining task performance requires more effort, the individual may perceive that he or she is working harder, even though performance levels are constant. Under these circumstances, we can expect that reports of subjective fatigue will be sensitive to increasing time-on-task whether or not performance is maintained. However, in the context of Davis' (1946) account, it is only when subjective fatigue is constant or decreasing with increasing time-on-task that one might expect a robust pattern of decreasing performance, such that the decrease in subjective fatigue would correspond to a withdrawal from the task.

These considerations further lead to the possibility that changes in subjective reports of cognitive fatigue may mediate time-on-task effects on performance through their effects on subsequent performance strategy. For some individuals, increasing feelings of cognitive fatigue associated with time-on-task may initiate changes in cognitive effort for the purpose of sustaining performance. For other individuals, feelings of cognitive fatigue may trigger a self-

protective orientation that results in reduced cognitive effort and declining performance as time-on-task increases. One aspect of the present study will involve examination of the possible association between subjective fatigue and strategic responses to increasing time-on-task.

Individual Differences in Susceptibility to Subjective Cognitive Fatigue

In contrast to research attention directed at situational and task influences on cognitive fatigue and performance, few studies have examined the potential influence of individual differences in non-ability traits on cognitive fatigue. As De Vries and Van Heck (2002) argued, transituational traits are also likely to influence feelings of subjective fatigue. That is, some individuals are likely to report subjective fatigue even in the absence of an explicit fatiguing event, and some individuals are likely to be more reactive to a fatiguing event than other individuals. Obvious candidate traits that may be correlated with transituational subjective fatigue are those that relate to neuroticism, anxiety, and introversion-extroversion, because of their associations with self-reports of baseline arousal and various reported physical symptoms and feelings (e.g., see Matthews et al., 1999; Gray, 1990; Revelle, 1993; Watson, Wiese, Vaidya, & Tellegen, 1999). Other related nonability traits that have been implicated as related to subjective reactions to performance in achievement/evaluation settings, such as attitudes toward learning, mastery, need for achievement, and competitiveness, may also be related to subjective fatigue in work and testing contexts, although there has been no comprehensive research to date on the trait correlates of short-term subjective fatigue in these contexts.

Trait Complex Assessment

To evaluate the role of a broad range of nonability traits on subjective cognitive fatigue, we use the trait complex methodology (e.g., see Ackerman & Heggestad, 1997; Armstrong, Day, McVay, & Rounds, 2008; Staggs, Larson, & Borgen, 2007; Sullivan & Hansen, 2004). This approach allows for a more parsimonious representation of key personality, interest, and motivational trait measures that may share substantial communality than an approach that considers each trait in isolation. Based on Snow's conceptualization of aptitude complexes (Snow, 1963), Ackerman (1996) introduced the trait-complex approach and has obtained empirical evidence suggesting that trait complexes represent constellations of traits that may be impeding or facilitative of learning and performance in specific domains, in that they represent constellations of traits that have a synergistic effect on the orientation of the individual toward or away from such situations. Within the framework, the choice of which trait complexes to assess generally relates to the specific aims of a particular investigation. To derive trait complex measures, a combination of top-down and bottom-up methods are used. First (top-down), a large battery of trait measures are selected that include markers for each of the trait complexes. Second (bottom-up), trait complexes are identified from a factor solution of trait measures. Finally, trait complex measures are created using unit-weighted *z*-score composites that have salient loadings on the underlying factors. The breadth of each trait-complex measure is then assessed by computing internal consistency reliability (α) values based on the constituent scales for each composite.

To date, several trait complexes have been shown to have positive or facilitative relations with intellectual performance and in short-to-medium term learning tasks, in that they are associated with an orientation *toward* and interest in tasks with intellectual content. Those trait complexes include Need for Achievement/ Mastery, Desire to Learn/Typical Intellectual Engagement. Other trait complexes have been associated with either an orientation *away* from tasks with intellectual content or an interest in activities that are not intellectual (e.g., social). Those trait complexes include Neuroticism/Anxiety, Extraversion, Extrinsic Goal Orientation, and Competitiveness/Other Goal Orientation (for examples and discussion, see Ackerman, 2003; Ackerman & Beier, 2006; Ackerman, Bowen, Beier, & Kanfer, 2001; Ackerman & Heggestad, 1997; Ackerman & Wolman, 2007).

Based on the Ackerman and Kanfer (2006) review of prior theory and research on putative determinants of cognitive fatigue, we expected that six broad trait complexes would show predictive validity for individual differences in subjective fatigue. Specifically, we expect that two facilitating trait complexes, namely Need for Achievement/Mastery and Desire to Learn/Typical Intellectual Engagement, will be negatively related to subjective cognitive fatigue; and that four impeding trait complexes, namely Neuroticism/Anxiety, Extraversion, Extrinsic Goal Orientation, and Competitiveness/Other Goal Orientation, will be positively related to subjective cognitive fatigue.

A Conceptual Model of Cognitive Fatigue

Extant theories of cognitive fatigue (e.g., Grandjean, 1968; Hockey, 1983), originating from the experimental psychology tradition, along with more general theories of attention and effort (e.g., Kahneman, 1973) relate mainly to performance effects, while theories of cognitive fatigue from the industrial/organizational psychology literature mainly address the nomological network of subjective fatigue (e.g., Åhsberg, 1998) and the effects of long-term work conditions on subjective fatigue and recovery (e.g., Sonnentag, 2003; Sonnentag & Zijlstra, 2006). Two basic findings from experimental studies on fatigue are that: (1) subjective fatigue increases with time-on-task when there are no opportunities for breaks or off-task activities, and (2) subjective fatigue is not system-specific. For example, Åhsberg (1998) reported that there are five major components of subjective fatigue, namely lack of energy, physical exertion, physical discomfort, lack of motivation, and sleepiness. With the exception of Schmidtke (1976), who posited an "emergency capacity" for use when effort demanded exceeds the effort that the individual intended to devote to the task, Kahneman's (1973) notion that an increase in arousal associated with task demands leads to an increase in available attentional capacity, and Hockey's (1997) concept of compensatory control (i.e., increases in effort allocations when the individual is under high workload), other approaches generally assume that performance will decline as the task becomes fatiguing, mainly because there is only one source of attentional/effort capacity, which is replenished only by rest or engagement on a different task.

For the current study, we have adopted a conceptual model of subjective cognitive fatigue that integrates Schmidtke's (1976) reserve capacity with a representation of subjective fatigue as a function

of the perceived depletion of cognitive effort resources as time-on-task proceeds. Although a full review of the historical and empirical basis for this model is beyond the scope of this article, several key elements of the representation warrant note. First, consistent with goal theories (for a review, see Kanfer, 1990), we predict that an individual's intentions and goal structures influence motivation, which in turn determines effort to be allocated to the task. However, we also stipulate that once the individual has made a commitment to a performance goal (e.g., see Heckhausen & Kuhl, 1985), increases in perceived task demands will result in an increase in task effort, up to the individual's level of effort that he or she is willing to expend on the task; a conceptualization that is also consistent with Hockey's concept of compensatory control. Also, similar to Baumeister's theory and empirical findings regarding self-regulation depletion (e.g., see Schmeichel, Vohs, & Baumeister, 2003), as time-on-task continues, the main source of effort available is depleted by effort allocated to the task (and by attention drawn by off-task distractions). In a situation of high-stakes work or testing, as time-on-task proceeds, individuals may *maintain* or *increase* performance by increasing allocated effort, reducing off-task distractions, and/or by allocating effort from reserve capacity.

In contrast to the varied effects on performance, and consistent with prior theory and research, the current model implies that subjective cognitive fatigue will increase with time-on-task as a function of the depletion of effort available and increase in off-task distractions. Because the relationship between motivation and subjective fatigue is one of mutuality, as subjective fatigue increases beyond a critical level, motivation for task performance is expected to diminish and will be associated with a decline in conscious allocation of effort as the individual withdraws from task performance (e.g., Davis, 1946; Lewin, 1935). In addition, the current approach also is consistent with dissociations among subjective cognitive fatigue and performance, especially when the individual employs reserve effort capacity. The only condition where subjective fatigue is expected to decline as time-on-task increases is when the individual *reduces* effort on the task. Otherwise, subjective fatigue is expected to show an increase, regardless of the stability or change in performance levels, and is expected to increase more rapidly when the individual allocates effort from his or her reserve capacity.

Study Overview

Using the conceptual representation described earlier as our heuristic foundation, we examined three basic questions about the relationship between time-on-task, cognitive fatigue, and performance in a cognitively demanding task. First, we evaluated the effects of time-on-task (a situational factor) on test performance. Second, we examined the effect of time-on-task on the pattern of self-reported cognitive fatigue over the course of testing. Third, we investigated the trait determinants of subjective cognitive fatigue. To answer these questions, we obtained pretest assessment of distal trait complexes by asking participants to complete an extensive battery of self-report measures at home, approximately one to two weeks prior to the first testing session. The questionnaire included trait measures of personality, affect, interests and motivation. Next, participants were administered an SAT (cognitive ability test) of three different time-on-task test lengths over the next three consecutive Saturday mornings. The three different test lengths were administered, in a counterbalanced, within-participants design. The test lengths were: Standard (standard SAT test length) – 3 hr, 45 min of testing time over a 4½-hr total session (including instructions and brief breaks); Short (1 hr shorter than Standard; 3½-hr total session), and Long (1 hr longer than Standard; 5½-hr total session). Subjective judgments, attitudes, and reactions were assessed prior to, during, and immediately following each testing session. These measures included subjective fatigue, negative and positive affect, positive motivation, confidence, and self-efficacy. Finally, archival data were obtained with the participants' SAT performance under high-stakes testing conditions, for comparison with the performance measures in the current study.

A comment about the procedure is appropriate at this point, vis à vis the extant mental or cognitive fatigue literature. In the extant literature, the traditional approach to assessing performance effects associated with fatigue is to administer the same task with parallel items over extended period of trials or sessions (e.g., see Arai, 1912). However, in the study of mental fatigue (as opposed to motor fatigue), as noted by multiple authors over the past 90 years (e.g., see discussion by Carmichael, Kennedy, & Mead, 1949), this approach is fraught with interpretational indeterminacies, mainly because there is no control group, and no way to separate warm-up effects or practice/learning effects from fatigue effects. Indeed, a large number of studies that purported to show increases in performance during extended task/testing sessions were attributable to practice effects (e.g., see discussion by Thorndike, 1926). The current procedure, by using a between-session test-length manipulation and a counterbalanced design, distributes any practice effects across the three test-length conditions, such that any differences between the conditions cannot be attributed to practice or learning effects.

Hypotheses

Arguably, the most critical applied issue with respect to cognitive fatigue and test length is whether SAT performance declines with longer testing sessions. Extant models of cognitive fatigue (e.g., Grandjean, 1968) indicate that if the amount of attention devoted to the task exceeds typical replenishment rates with no opportunity for recovery (e.g., sleep, extended breaks, off-task activities, etc.), performance will decline as time-on-task increases. In contrast, if there is a reserve capacity that can be tapped as the main source of attentional effort is depleted, performance may remain stable over an extended period of time or even increase in response the perceived task demands. With the assumption that motivation for successful task performance is relatively constant across a test session, two competing predictions are as follows:

> *Hypothesis 1A:* As total session time increases, SAT performance will decline (i.e., performance will be best in the Short test session and worst in the Long test session).

> *Hypothesis 1B:* SAT performance will not change or will increase as a function of total testing time. Note that with a relatively large sample size and within-participant design, the power to detect even a small effect ($f = 0.10$) exceeds .99 (G∗Power3; Faul, Erdfelder, Lang, & Buchner, 2007).

The second set of hypotheses pertains to the effects of time-on-task on subjective cognitive fatigue. Because subjective fatigue is expected to increase as time-on-task increases, except when the individual withdraws from the task, we expect that:

Hypothesis 2A: Subjective fatigue will increase as time-on-task increases in each testing session.

In addition, the proposed representation suggests that cognitive fatigue will increase as a function of resource depletion, with greater time-on-task yielding higher levels of self-reported cognitive fatigue. Thus, we predict that:

Hypothesis 2B: Subjective fatigue will be highest in the final hour of testing in the Long test condition, next highest in the final hour of testing in the Standard-test condition, and lowest in the final hour of testing in the Short-test condition.

The third set of hypotheses pertain to the conceptualization of subjective fatigue as partly determined by stable person traits and partly determined by the test situation. From a trait perspective, we expected that individual differences in subjective fatigue would be relatively stable within a test session, across test sessions, and would share substantial commonality with both distal trait complexes and proximal affective and motivational reactions to the test situation. Thus:

Hypothesis 3: There will be substantial commonality among subjective fatigue measures before testing, during testing, and after testing within and between each test-length condition.

Hypothesis 4: Individual differences in subjective fatigue before, during, and at the end of testing will be substantially positively related to distal trait complexes identified as impeding of learning and cognitive performance and negatively related to distal trait complexes identified as facilitative of learning and cognitive performance.

Hypothesis 5: Subjective cognitive fatigue will share substantial common variance with self-efficacy and related judgments (e.g., motivation for task performance, confidence in task performance) and proximal affective self-reports (i.e., negative and positive affect).

In concert with the strategic differences in reactions to cognitively fatiguing conditions described by Davis (1946), our approach indicates an association between subjective fatigue and motivation. Specifically, we anticipated that individuals who reported decreasing effort levels during the test session would report having higher levels of cognitive fatigue, while individuals who reported stable or increasing levels of effort during the test session would be likely to report lower levels of cognitive fatigue. Thus:

Hypothesis 6: Individuals who retrospectively report decreasing effort across a test session will have higher levels of subjective fatigue than those who report constant or increasing levels of effort during the session.

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 12 participants were eliminated from analysis. Six of the participants either failed to follow instructions or fell asleep and had to be awakened repeatedly during testing. The other six students exerted minimal effort during testing (with average scores approximately 200–400 points, based on Verbal[1] + Quantitative composites, below their archival record scores across all three sessions of the study). Among the remaining 239 participants, 124 were men and 115 were women. Mean age of the sample was 18.3 (SD = .53) years, and participants ranged from ages 17 to 20 years. 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*.

The At-Home Questionnaire (AHQ)

The AHQ included measures of traits to be used in the derivation of trait complex composites.

Trait assessments. The AHQ packet included several self-report scales designed to assess affective and conative traits that were expected to be related to subjective reports of fatigue and other attitudes and emotions during testing, along with test performance. The scales were designed to provide robust markers for a the following trait complex factors: Need for Achievement/Mastery, Desire to Learn/Typical Intellectual Engagement, Neuroticism/Anxiety, Extroversion, Extrinsic Goals, and Competitiveness/Other-Oriented Goals (e.g., see Ackerman et al., 2001; Ackerman & Wolman, 2007). A list of all of the trait-based scales from the AHQ is provided in the Appendix, along with the number of items, mean, SD, and internal consistency reliability (α) in Table A1.

Additional scales were administered, but are not presented here, because of space limitations.

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 1. As indicated in Table 1, the first 2½ hr 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

[1] With the introduction of the new version of the SAT in March, 2005, the Verbal section was renamed to "Critical Reading" and the Math section was renamed to "Mathematics." For consistency with the extant literature and common usage, we have retained "Verbal" and "Math" terms for these sections in this article.

Table 1
Layout of Test Sections by Condition

	Long condition	Time (hr:min:s) 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-min SAT test segments shown in italics.

verbal, math, or writing sections. Equivalent 50-min 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.

SAT test scoring 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 locally, using a (number correct) − (.25 ∗ number wrong) formula.

Subjective Reports During Testing

Pretest subjective reports were obtained at the start of each test session (before any SAT test items were administered). Interim questionnaires (with items identical to the pretest questionnaire) were administered at 2½ hr into each session. Additional retests of the interim questionnaires were administered after 3½ hr of testing in the Standard and Long conditions, and after 4½ hr 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 hr in the Short condition, 4½ hr in the Standard condition, and 5½ hr 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, and after the final section of testing. The questionnaire included: scales of Subjective Fatigue (12 items, e.g., "worn out," "exhausted"; α = .91), Negative Affect (12 items; e.g., "uneasy," "tense"; α = .87); Positive Affect (7 items; e.g., "enthusiastic," "excited"; α = .89). These items were provided with a 5-point scale (1 = *very slightly or not at all*, 2 = *a little*, 3 = *moderately*, 4 = *quite a bit*; 5 = *extremely*). Additional scales included 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"; α = .88) 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]; α = .87). These scales were provided with a 6-point Likert-type scale (1 = *strongly disagree*, 2 = *moderately disagree*; 3 = *slightly disagree*; 4 = *slightly agree*, 5 = *moderately agree*; 6 = *strongly agree*).

Additional Questions at End of Each Session

At the end of the posttest for each session, an additional set of questions was included. These questions pertained to the individual's report of increasing or decreasing effort during the course of the session.

Archival Data

Following completion of the study, archival records were obtained of the participants' SAT scores, from the tests taken under high-stakes testing situation.

Procedure

Testing took place in a large classroom with from 48 to 84 participants at a time, on three consecutive Saturday mornings. In all conditions, participants were instructed to arrive a few min before the official 8 a.m. start of testing. (On the first day of testing for each group, consent and records 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 prerecorded 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 average SAT equivalence scores (across the three test sessions) met or exceeded the archival SAT test scores (taken under high-stakes conditions). Participants were also told that the five participants with the largest increase in scores in comparison to the archival scores would receive an additional $100 bonus. The experimental layout for each of the three conditions is shown in Table 1. At the end of the third and final session, participants were debriefed and compensated $150 for their participation. Bonuses were sent to the qualifying participants after the archival records were obtained.

Results

Performance Scores

Two sets of scores represent the critical performance 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 is 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 was administered (one 20 min Verbal test, one 20 min Math test, and one 10 min multiple-choice Writing test). Essay tests were not scored, partly because they occurred during the initial part of each test (and thus were less likely to be influenced by fatigue), and also for logistical reasons (given the unavailability of trained graders for these tests).

Test of counterbalancing/order effects. A comparison was made of average SAT equivalence scores 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 in all subsequent analyses.

Comparison with performance under high-stakes testing conditions. The first analysis for the SAT equivalence scores was a comparison with College Board (CB) records, to evaluate the overall difference between the high-stakes testing situation (CB records) and the lower-stakes testing of the laboratory experiment. SAT Equivalence performance across the three sessions yielded a mean of 1229, and $SD = 164.64$. From the participants with matching CB records ($N = 222$), SAT performance under high-stakes testing was $M = 1281$, $SD = 159.12$. A paired t test indicated that mean performance under high stakes testing was 52.09 points higher than under the laboratory testing conditions, which was significant, $t(221) = 9.14$, $p < .01$, $d = .32$. However, 48% of the participants obtained average SAT equivalence scores that equaled or exceeded their CB record scores. In comparison to the archival SAT scores, the range of differences was from a drop of −253 points to an increase of 146 points. (Of the participants that had lower scores in the study, two thirds of the participants had scores that were less than or equal to a 100 point difference—a threshold that represents .62 *SD* units, or a medium-sized effect.) 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 (this might especially be the situation for participants whose archival SAT score were near the maximum score of 1600); (b) the conditions of testing had a negative effect on their performance; (c) the decline in performance was the result of regression to the mean (especially for those participants with very high archival SAT test scores).

SAT equivalence scores by session. An analysis of variance (ANOVA) performed on overall SAT equivalence scores to Test Hypotheses 1A and 1B pertaining to the effects of total time—on-task on performance indicated a significant effect of session, $F(2, 364) = 6.27$, $p < .01$. That is, the longer sessions yielded *higher* mean performance scores on the SAT than the shorter sessions, with the Standard length session showing an increase of 13 points over the Short session, and the Long sessions indicating a 15 point average increase over the Standard session. For the Short session, SAT equivalence scores were $M = 1{,}209$, $SD = 173$; for the Standard session, $M = 1{,}222$, $SD = 174$; and for the Long session $M = 1{,}237$, $SD = 177$. 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, shown in Figure 1. For the final 50 min of each session, scores were similar (Short $M = 34.06$, $SD = 8.85$; Standard $M = 34.48$, $SD = 9.10$; Long $M = 34.34$, $sd = 9.35$). An ANOVA indicated no significant

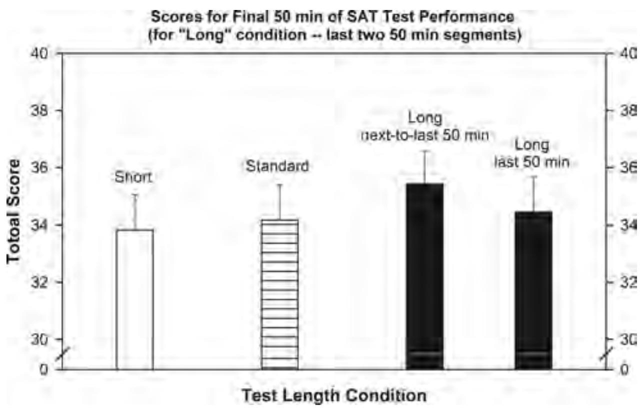

Figure 1. SAT test performance (number correct minus a fraction for number wrong) by test-length condition for the final 50-min section (Verbal [20 min] Quantitative [20 min] + Writing [10 min] test components). For the Long test length condition, both the last 50 min section and a parallel penultimate 50 min section are shown. Error bars indicate 95% confidence intervals for the means.

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 the Short, Standard, and the final 50 min of the Long sessions, respectively. The mean scores and 95% confidence intervals are shown in Figure 1.

The results of analyses on test performance provide no support for Hypothesis 1A (a decline in test performance in the longer test sessions), but rather we did find support for Hypothesis 1B (no change or an increase in test scores as a function of test length). Not only did mean SAT test scores fail to show a decline as a function of test length, but mean scores actually significantly *increased* from Short to Standard test length conditions, and significantly *increased* from Standard to Long test length conditions.

Summary. The use of nearly identical overall procedures as used in standardized SAT testing (including day of the week, time of day, instructions, and so on), along with the provision of cash bonuses for participants who met or exceeded their archival SAT records scores appears to have created a testing situation that was largely congruent with the typical SAT testing experience, even though the participants were already enrolled in postsecondary institutions at the time of 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.

In contrast to Hypothesis 1A positing a decline in test performance as length of test increased from 3½-hr to 5½-hr total time, the results indicated that average SAT equivalence scores actually *increased* with increasing test session length—consistent with Hypothesis 1B. Because this experiment used a within-participants 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 SAT test, repeated after approximately each hour of testing, and 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 subjective 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, subjective fatigue increased with increasing time-on-task, providing support for Hypothesis 2A (subjective fatigue will increase as time-on-task increases in each condition).

Across conditions, a comparison of subjective fatigue prior to the last hour of testing, showed a significant increase in subjective fatigue from the Short to the Standard to the Long test condition ($M = 2.89$, 3.02, and 3.17, respectively), $F(2, 356) = 8.25$, $p < .01$, $MSE = .47$. These results provide support for Hypothesis 2B, that subjective fatigue in the last hour of testing would be highest for the Long test condition, followed by the Standard test condition, and then the Short test condition. In terms of effect size, the difference between posttask subjective fatigue between the Short and Standard conditions was $d = .21$, between Standard and Long test conditions was $d = .09$, and between the Short and Long test conditions was $d = .32$. Although these are not large effects in terms of Cohen's (1988) rules-of-thumb, a more salient contrast might be made when comparing the percent of participants reporting substantial levels of subjective fatigue in the three conditions. At posttest, 27%, 32%, and 36% of the participants reported either "Quite a Bit" of fatigue or being "Extremely" fatigued, in the Short, Standard, and Long test length conditions, respectively. A Cochran Q test of these differences was significant ($Q = 8.09$, $p < .05$).

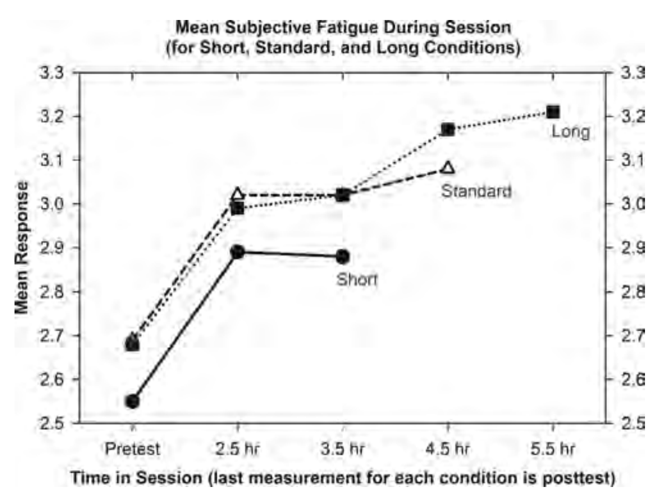

Figure 2. Mean subjective fatigue scale scores for pretest, interim, and posttest administrations, by test condition.

Although not hypothesized, a few across-condition comparisons related to self-reported cognitive fatigue also warrant note. First, pretest subjective fatigue for the Short condition was significantly lower than the pretest subjective fatigue for the Standard condition, $t(185) = -2.71$, $p < .01$, $d = -.16$, but only marginally lower, in comparison to the Long condition ($t(186) = -1.74$, *ns*, $d = -.12$). Immediate posttest assessments of subjective fatigue were also lower for the Short condition, compared to the Standard condition, $t(185) = -3.91$, $p < .01$, $d = -.21$ and the Long condition, $t(186) = -5.10$, $p < .01$; $d = -.32$. 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*, $d = -.09$). Finally, there was no noticeable decline (recovery) in subjective fatigue from the last interim administration to the assessment immediately after testing.

Individual Differences in Subjective Fatigue

The intercorrelations among the subjective fatigue measures in the three test length conditions are shown in Table 2. To evaluate Hypothesis 3, we examined the correlations between subjective fatigue measures administered just prior to the last hour of testing in each condition. The correlations between these measures were $r = .62$ (Short vs. Standard), $r = .47$ (Short vs. Long), and $r = .61$ (Standard vs. Long), all of which were significant at $p < .01$. Squaring the correlations provides an index of common variance, yielding a value of 22% to 38% shared variance across the three conditions. Thus, there was substantial commonality among perceptions of subjective fatigue across the three sessions that were completed in three different weeks, and occurring at 2½, 3½, and 4½ hr into the respective sessions. In general, the hypothesis stated there would be substantial common variance among subjective fatigue measures within and across test sessions. As shown in Table 2, all correlations are positive and significant. As expected for any multioccasion assessments, the repetitions of subjective fatigue assessments within each condition show a simplex-like pattern (e.g., see Ackerman, 1987; Humphreys, 1960), 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. Even for the most distant pair (Short condition pretest with Long condition posttest), the correlation ($r = .35$) remained positive and significant. Overall, the pattern of results obtained provide support for Hypothesis 3.

In addition, the pattern of correlations shown in Table 2 indicates a high degree of commonality among the pretest subjective fatigue measures across the conditions ($r = .68$ for Short vs. Standard, .53 for Short vs. Long; and .58 for Standard vs. Long). That is, in addition to the predictive value of pretest subjective fatigue for posttest subjective fatigue in each test condition, pretest subjective fatigue shared 28% to 46% of the variance with other session pretest subjective fatigue conditions. Individuals who expressed higher levels of subjective fatigue prior to the Short test condition, also had a substantial tendency to express higher levels of subjective fatigue in the longer test conditions.

Trait Complexes

The distal trait measures were subjected to a 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 (i.e., loadings above .30 on a single factor), 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 3. 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 Table 3.

Table 2
Means, SDs, and Cross-correlations Among Subjective Fatigue Measures

	Mean	*SD*	1	2	3	4	5	6	7	8	9	10	11
1. Short pretest	2.55	.92											
2. Short interim	2.89	1.04	**.74**										
3. Short posttest	2.88	.99	**.70**	**.86**									
4. Standard pretest	2.69	.94	***.68***	.54	.57								
5. Standard interim 1	3.02	1.06	.48	.57	.60	**.74**							
6. Standard interim 2	3.02	1.06	.54	.60	.69	**.72**	**.86**						
7. Standard posttest	3.08	1.02	.52	.62	.72[a]	**.68**	**.80**	**.90**					
8. Long pretest	2.68	.91	***.53***	.46	.46	***.58***	.52	.53	.49				
9. Long interim 1	2.99	.98	.37	.52	.57	.44	.58	.59	.55	**.67**			
10. Long interim 2	3.02	1.00	.36	.53	.61	.45	.60	.64	.64	**.61**	**.83**		
11. Long interim 3	3.17	1.07	.34	.47	.58	.41	.58	.61	.65	**.47**	**.69**	**.87**	
12. Long posttest	3.21	1.04	.35	.47	.62[a]	.38	.55	.64	.69[a]	**.48**	**.66**	**.82**	**.90**

Note. Boldface indicates within condition correlations. Italics indicate correlations between pretest measures. $N = 185$ to 209. Correlations larger than $r = .18$, significant $p < .01$.
[a] Correlations between posttest measures.

Table 3
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**

Note. $N = 239$, $df = 237$. N = neuroticism; nAch = need for achievement; DTL = desire to learn; TIE = typical intellectual engagement.
* $p < .05$. ** $p < .01$.

Correlations Between Trait Complexes and Subjective Fatigue During SAT Testing

Hypothesis 4 stated that distal trait complexes would share commonality with subjective fatigue measures taken before and after each test session. Table 4 provides the correlations for each of the administrations of subjective fatigue. In support of Hypothesis 4, the trait complexes show a robust pattern of significant and salient correlations with the self-report measures obtained before, during and after the SAT testing. Although not all of the trait complexes were significantly correlated with subjective fatigue measures, subjective fatigue showed significant positive correlations with the N/Anxiety trait complex, and negative correlations with both the nAch/Mastery and DTL/TIE.

Table 4
Correlations Between Trait Complexes and Subjective Fatigue During Testing

Subjective fatigue	nAch/Mastery	DTL/TIE	N/Anxiety	Extroversion	Extrinsic Goals	Competitiveness
Short pretest	−.14*	−.23**	.30**	.06	−.06	.15*
Short interim	−.16*	−.18*	.34**	.07	−.03	.12
Short posttest	−.12	−.20**	.34**	.03	−.04	.15*
Standard pretest	−.29**	−.30**	.31**	.05	−.12	.11
Standard interim 1	−.22**	−.22**	.35**	−.00	−.02	.06
Standard interim 2	−.25**	−.21**	.35**	.00	−.04	.15*
Standard posttest	−.20**	−.15*	.37**	−.01	−.02	.18**
Long pretest	−.27**	−.21**	.21**	.20**	−.09	.02
Long interim 1	−.22**	−.21**	.29**	.09	−.01	.02
Long interim 2	−.17**	−.20**	.28**	.10	−.04	.06
Long interim 3	−.21**	−.17*	.30**	−.00	−.03	.04
Long posttest	−.19**	−.11	.30**	−.00	−.01	.04

Note. $N = 185$ to 209. N = neuroticism; nAch = need for achievement; DTL = desire to learn; TIE = typical intellectual engagement.
Correlations larger than * $p < .05$; ** $p < .01$.

Even though individual trait complexes (especially nAch/Mastery, DTL/TIE, and N/Anxiety) 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 5. For the first model, the regression findings indicate that trait complexes from the AHQ account for from 22% to 30% of pretest subjective fatigue, and from 22% to 24% of the variance in posttest fatigue across the three testing conditions.[2] 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 32% 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, account 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 16% 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 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 variance accounted for by trait complexes/self-concept/self-estimates of ability, Short condition posttest fatigue, and pretest fatigue was 66% in the Standard testing condition, and 50% in the Long testing condition. In all cases, adding SAT performance to the prediction of subjective fatigue did not result in any incremental predictive validity, after the trait complex variables and other subjective fatigue measures were entered into the equation. This result supports the notion that consideration of individual differences in actual test performance provided no useful incremental information about subjective fatigue.

Summary

Taken together, the results show significant relationships between distal trait complexes and individual differences in subjective fatigue, before, during, and after testing provide support for Hypothesis 4. Approximately 20% to 30% of the variance in pretest and posttest subjective fatigue was accounted for by distal trait complex measures assessed at a time different from the actual testing day. An additional 11% to 15% of the variance in pretest fatigue is accounted for by posttest subjective fatigue in a Short testing session. In comparison, 14% to 30% of the variance in posttest fatigue is accounted for by pretest fatigue on the testing day.

Proximal Correlates of Subjective Fatigue

In addition to the measurement of subjective fatigue, before, during, and immediately after each test session, assessments were made of self-reported negative affect, positive affect, positive motivation, confidence in upcoming test performance, and self-efficacy. Aggregate scores across the assessments and test sessions were correlated, and are shown in Table 6. Subjective fatigue showed significant correlations with each of these other self-report measures, most notably with negative affect ($r = .58$) and with confidence for test performance ($r = .70$). A multiple correlation with subjective fatigue as the criterion variable and the other five self-report variables as predictors, was $R = .83$, $p < .01$, indicating that there was 69% shared variance among these measures. Although there is a multicollinearity among these measures, multiple regression analyses indicated that confidence in test performance had the largest contribution to the prediction of subjective fatigue ($\beta = -.62$), followed by negative affect ($\beta = .38$), with smaller, but significant contributions from self-efficacy and positive affect measures. The pattern of substantial common variance observed between subjective fatigue and self-efficacy, affective states, and motivation for task performance provides support for Hypothesis 5.

Test-taking Strategy and Subjective Fatigue

Participants were also asked at the end of each session to select one from among four effort strategies that best described their approach during the session. The four strategies were “kept my effort at a constant level,” “increased my effort,” “decreased my effort,” and “first I increased my effort, then later I decreased my effort.” In the Short testing condition, 16% of the participants reported using an increased effort strategy, 61% reported using a constant level effort strategy, 13% reported a decreased effort strategy, and 10% reported increasing, then decreasing their 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. Overall, the pattern of results show remarkable stability in the distribution of self-reported effort strategies used across conditions. Specifically, approximately one half of the participants reported using a constant effort strategy in each of the test length conditions, with the remaining percent of participants relatively evenly spread across the other three strategies.

To Test Hypothesis 6—that participants who indicated a reduction of effort would have higher levels of subjective fatigue, we computed mean subjective fatigue scores, conditioned on the strategy indicated for the three test length conditions. As can be seen in Figure 3, participants who reported either decreasing their

[2] Prior to each session, we asked the participants to indicate what time they went to bed and what time they awakened on the day of the test session. To support the point that pretest fatigue is something more than how “tired” the participants were, we note that although correlations between amount of sleep and pretest subjective fatigue were significant in each condition (*r* .19, .25, and .28, all *p* .05 for the Short, Standard and Long test-length conditions), in each case the amount of sleep accounted for a much smaller amount of variance in individual differences in pretest subjective fatigue than did the trait-complex measures.

Table 5
Summary of Hierarchical Regressions for Predicting Subjective Fatigue During Testing (With SAT Performance at Last Step)

Model 1. Trait complexes in Step 1. Pretest subjective fatigue in Step 2

	Step 1: Trait complexes	Step 2: Pretest fatigue	Final step: SAT performance
Short pretest			
R^2 to add	.20**		
Total R^2	.20**		
Short posttest			
R^2 to add	.22**	.32**	.00
Total R^2	.22**	.54**	.54**
Standard pretest			
R^2 to add	.27**		
Total R^2	.27**		
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**	.00
Total R^2	.22**	.36**	.36**

Model 2. Step 1: trait complexes. Step 2: posttest subjective fatigue in short condition; Step 3: pretest subjective fatigue

	Step 1: Trait complexes	Step 2: Short condition posttest fatigue	Step 3: In-session pretest fatigue	Final step: SAT performance
Standard pretest				
R^2 to add	.27**	.16**		.00
Total R^2	.27**	.43**		.43**
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**	.00
Total R^2	.22**	.46**	.50**	.50**

Model 1: Step 1 df = 6,153; Step 2 df = 1,152; Step 3 df = 1, 151. Model 2: Step 1 df = 6,153; Step 2 df = 1,152; Step 3 df = 1,151; Step 4 df 1,150.
** $p < .01$.

effort, or decreasing their effort after an increase in effort earlier in the sessions, also reported greater subjective fatigue, in comparison to individuals who reported keeping a constant level of effort or increasing their effort over the course of session. Significant differences were found for each of the test sessions, $F(3, 195) = 11.01$ for the Short test condition; $F(3, 198) = 4.51$ for the Standard test condition; and $F(3, 200) = 5.70$ for the Long test condition, respectively; all $p < .01$. These results provide support for Hypothesis 6, and indicate that increases in subjective fatigue were associated with reported strategies that involved reduction of effort during the test sessions.

By comparison, the effects of reported effort strategy were attenuated for performance. An analysis of last 50 min SAT test performance by reported strategy showed significant differences among these groups *only* 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. That is, only individuals who reported decreasing effort during the Short test condition obtained significantly lower performance, compared to participants who kept constant or increasing levels of effort during the session.

Table 6
Correlations Between Composite Self-report Measures (Across Short, Standard, and Long Conditions)

Composite self-report measures	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**

** $p < .01$.

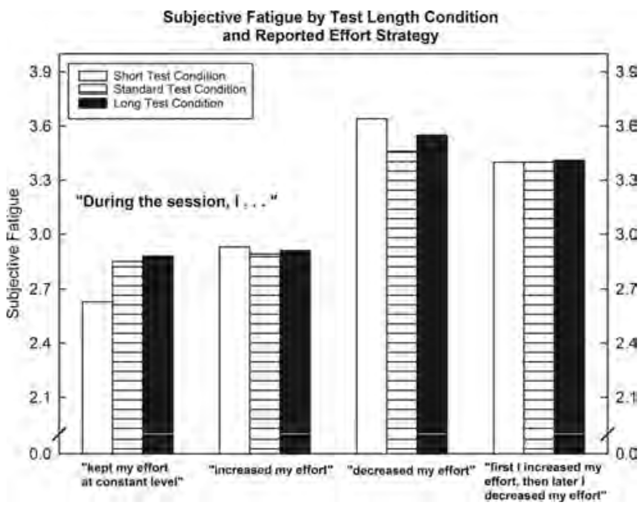

Figure 3. Mean subjective fatigue by test length condition and self-reported effort strategy.

In addition to the within-condition analyses, we also classified participants who showed an identical strategy pattern across all tests versus those who reported any switch in strategy. One hundred eighty-two of the participants reported a switch in strategy across the conditions, while 57 reported identical strategy patterns. Of the 57, there was only a substantial number ($N = 47$) in the category who reported constant level of effort throughout the sessions. A comparison of those participants with the participants who changed strategy indicated significantly lower SAT equivalence scores for those who changed strategies in the Short condition, $t(191) = 1.98$, $p < .05$, $d = .35$ and the Standard condition, $t(197) = 2.67$, $p < .01$, $d = .48$, but a nonsignificant difference for the Long condition ($t(199) = 1.69$, *ns*., $d = .29$). Overall, these results support the notion that participants who reported maintaining a constant level of effort within and across test-length conditions tended to perform better overall, in comparison to those participants who reported reducing their effort in any of the test-length conditions.

Discussion

Although there was colloquial and mainstream media support for the hypothesis that SAT performance in the longer test-length sessions would be lower than that of the short test-length sessions (Hypothesis 1A), our interpretation of the extant literature was relatively agnostic about finding such differences. In fact, there was both theoretical and empirical support for the proposition that performance would remain constant or increase across the range of test-length sessions employed in this study (Hypothesis 1B). The results of this investigation showed clearly that performance was not adversely affected in the longer test-length sessions. In contrast to performance effects, there was little, if any, controversy about increases in mean levels of subjective fatigue as test-session length increased (Hypothesis 2). All extant theories and nearly all prior empirical literature supported this hypothesis, and the data from the current study were consistent with this hypothesis.

In contrast to the hypotheses about mean effects, Hypotheses 3–5 pertained to individual differences in subjective fatigue, an area that has had very little prior literature, and none in the kind of experimental paradigm under consideration. The trait versus situation controversy that raged in the personality and social psychology literature in the 1970s provides one backdrop for the current findings. In this experiment, we used a person and situation interaction approach to examine the determinants of cognitive fatigue in a cognitively demanding test like those routinely used in selection settings. Consistent with the proposed framework for cognitive fatigue, the results show that individual differences in trait complexes significantly predict subjective cognitive fatigue, and that subjective cognitive fatigue increases with time-on-task, with the qualification that only three of the six trait complexes had consistently substantial correlations with subjective fatigue measures. Perhaps the most important aspect of this investigation was the dissociation between performance scores and subjective fatigue self reports. The pattern of increased subjective cognitive fatigue and increased performance over time-on-task is particularly interesting because it provides empirical support for the notion that the perception of fatigue may not be simply epiphenomenal or an indicator of reduced performance but rather may reflect a functional process associated with level or increased effort. Further research is needed to identify the self-regulatory processes that govern the association between subjective fatigue and performance in test or work settings.

The dissociation between subjective fatigue on the one hand and performance on the other hand suggests that one needs to look beyond task-based behaviors for the consequences of high levels of subjective fatigue. Anticipations of subjective fatigue may lead some individuals to avoid tasks altogether. In the case of the SAT and similar tests, such individuals might avoid testing or retesting because of the anticipated fatigue, where other individuals might seek the opportunities associated with taking the test multiple times. In addition, individuals who experience high levels of subjective fatigue may misattribute their performance levels to their subjective states rather than to their underlying skills and abilities. Subjective attributions of fatigue, along with other related constructs, such as self-efficacy, may ultimately relate more to decisions of whether to engage the task at all, even though they do not substantially relate to performance, *if the individual decides to engage the task*.

Two other important findings from this study pertain to the influence of person characteristics on cognitive fatigue. First, using a trait complex approach we found that personality/interest/motivation trait complexes accounted for significant levels of subjective fatigue across all test length conditions, and throughout each testing session. 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 reported higher levels of DTL/TIE or nAch/Mastery reported lower levels of subjective fatigue throughout the study. In the aggregate, distal trait complex predictors accounted for between 22% and 30% of the variance in subjective fatigue measures. These findings provide strong evidence for the potential value of the trait-complex approach for investigating person influences associated with attenuated and heightened levels of subjective cognitive fatigue.

Results of this study also reveal that self-reported cognitive fatigue prior to the onset of testing may represent another important another person characteristic. Specifically, pretest reports of cognitive fatigue showed significant predictive validity for subjective fatigue measures taken at the end of testing. We found that

pretest subjective fatigue accounted for nearly half the variance in posttest fatigue in the Short and Standard test length conditions. Although the amount of posttest fatigue variance accounted for by pretest fatigue was somewhat less in the Long test length condition (23%), these results suggest that reported fatigue prior to testing plays an important role in subsequent experiences of fatigue in extended testing. The influence of person characteristics, including pretest cognitive fatigue and individual differences in trait complexes, on posttest reports of subjective fatigue is particularly noteworthy if one considers that between 36 and 55% of the variance in posttest subject fatigue could be accounted for by measures administered *prior* to any current engagement with the SAT test. Nonetheless, we did find support for Hypothesis 6, that is, that individuals who retrospectively reported a reduction in effort across the test session also reported higher levels of subjective fatigue *and* significant reductions in performance, indicating that broad strategic approaches to fatiguing situations may ultimately have some influence on an individual's cognitive effectiveness.

Limitations

The main limitation of this study was the relatively diminished stake in the outcome of the test performance. Although we attempted to duplicate the SAT test experience as closely as possible (e.g., using the same administrative procedures as used in the operational testing environment, including start times and breaks), the participants in the study had already matriculated to university study, and the results of the test had no impact on their future educational placement. Some regression-to-the-mean could be expected from a sample with above average SAT scores; however, about half (48%) of the participants met or exceeded the scores taken a year earlier under high-stakes testing. It is likely that many participants did not have either the same high level of motivation for performance nor the degree of anxiety about their performance in these test sessions than they did when they completed the SAT under high-stakes conditions. Similarly, the participants were not likely to study or engage in other preparatory activities for these test sessions.

The SAT test, while generally considered to be a robust test of quantitative and verbal abilities, arguably does not contain the kind of achievement content that would be representative of many professional certification tests, where domain knowledge plays a more central role. To the degree that domain knowledge tests depend less on fluid intellectual abilities and more on crystallized intellectual abilities (e.g., see Ackerman, 2000), such tests may be less susceptible to high levels of continuous cognitive effort (which is seen as generally analogous to fluid intellectual abilities—see Kanfer & Ackerman, 1989). Thus, it may be that cognitive fatigue represents a less critical factor in some professional certification tests, especially those taken by individuals with substantial professional experience.

Finally, as with most studies that involve assessment of subjective fatigue during task performance, reactivity may have influenced the results. Asking participants about their fatigue may have induced some participants to *feel* more fatigued as time-on-task proceeded. Although reactivity effects to self-report measures during cognitive testing have not typically been reported in the literature, a between-participants design with groups that only receive a single assessment of subjective fatigue at designated assessment points during task performance would be needed to address this possibility.

Conclusions

The increasing use of cognitive ability tests for educational and organizational selection purposes, along with the growing number of jobs that require prolonged periods of high cognitive effort has revitalized research interest in cognitive fatigue, in particular the relationships between time-on-task, subjective fatigue, and performance. Using the SAT context, we found that although self-reports of cognitive fatigue showed the expected increases within session and between test length conditions, mean levels of test performance modestly increased - rather than declined - in the longer test length conditions. These findings, demonstrating the dissociation between cognitive fatigue and performance, suggest that greater research attention should be given to understanding the psychological processes that regulate allocations of effort to task performance in the presence of feelings of cognitive fatigue.

One potentially fruitful direction for future research pertains to individual differences in nonability traits and their possible role in strategy choice and self-regulation of effort. Whether traits affect cognitive fatigue through action processes, or simply represent transsituational consistency in the self-report of fatigue remains an empirical question. Pragmatically, the finding that 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 suggests that self-reports of fatigue may be associated more with the broader "testing" context than longer test times.

Our findings do not suggest, however, that self-reports of cognitive fatigue are merely epiphenomena. Longer testing times did lead to increases in reports of subjective fatigue that did not recover immediately at the end of the testing session. Although Sonnentag and her colleagues (e.g., Sonnentag, Binnewies, & Mojza, 2008; Sonnentag & Zijlstra, 2006) have shown that leisure activities may enhance recovery from work-related fatigue, little is known about the effect of time-on-task on rate of cognitive fatigue recovery or the impact of interventions that facilitate recovery. In addition, given that we found proportionately more of the variance in posttest fatigue was predicted from stable traits than test length condition, rate of recovery from cognitive fatigue is likely to be a function of both the situation and the person.

From the test-taker's perspective, our findings present a conundrum. Test takers are likely to experience cognitive fatigue during longer tests, especially as time-on-task proceeds, and feelings of cognitive fatigue throughout testing are likely to be more pronounced among individuals high in N/Anxiety, low on DTL/TIE, and low on nAch/Mastery. Yet the experience of fatigue during testing does not appear to be, in and of itself, detrimental to test performance. Although reports of elevated cognitive fatigue were not associated with lower levels of performance (except for the Short condition), higher reports of cognitive fatigue were associated with reported reduction of effort during the session. Taken together, this pattern suggests that interventions that explicitly train new affect-action bonds—for individuals to respond to feelings of cognitive fatigue with higher, rather than lower levels of cognitive effort—may help to preclude premature withdrawal of task effort and associated lower levels of performance.

The findings obtained also have implications for broader issues related to cognitive fatigue, and for the role of trait-complex determinants of subjective experiences in tasks requiring extended expen-

ditures of cognitive effort. Although these results are based on a relatively small sample of time (12.5 hr across three sessions), and a sample of young adults, it may be the case that a larger proportion of the variance in subjective reports of subjective fatigue during task performance can be attributable to stable trait complexes, than by task requirements. For example, individual differences in affective and conative traits may influence subjective feelings that occur in the school or the workplace, which in turn affect how the individuals react to the job requirements (e.g., the expression of strain or job tension), rather than through a direct effect on task performance (Zellars, Perrewé, Hochwarter, & Anderson, 2006). Our findings provide further support for the notion that the critical pathway by which nonability traits influence cognitively demanding performance may be indirect, through the effects of traits on how the individual reacts to workplace conditions and the effect of those reactions on effort allocation and performance strategy. Additional empirical research is needed to examine whether these findings scale up to longer periods of time, such as across weeks or months on the job.

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. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence, 22,* 229–259.

Ackerman, P. L. (1997). Personality, self-concept, interests, and intelligence: Which construct doesn't fit? *Journal of Personality, 65,* 171–204.

Ackerman, P. L. (2000). Domain-specific knowledge as the "dark matter" of adult intelligence: Gf/gc, personality and interest correlates. *Journal of Gerontology: Psychological Sciences, 55*B, P69–P84.

Ackerman, P. L. (2003). Aptitude complexes and trait complexes. *Educational Psychologist, 38,* 85–93.

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. Atlanta, GA: 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.

Åhsberg, E. (1998). *Perceived fatigue related to work.* University of Stockholm/National Institute for Working Life (p. 19). Stockholm: National Institute for Working Life.

Arai, T. (1912). Mental fatigue. *Contributions to Education* (Vol. 54, pp. 1–115). New York: Teachers College, Columbia University.

Armstrong, P. I., Day, S. X., McVay, J. P., & Rounds, J. (2008). Holland's RIASEC model as an integrative framework for individual differences. *Journal of Counseling Psychology, 55,* 1–18.

Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology, 43,* 695–716.

Bartley, S. H., & Chute, E. (1947). *Fatigue and impairment in man.* New York: McGraw-Hill Book Company.

Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2005). Effects of mental fatigue on attention: An ERP study. *Cognitive Brain Research, 25,* 107–116.

Carmichael, L., & Dearborn, W. F. (1947). *Reading and visual fatigue.* Cambridge, MA: The Riverside Press.

Carmichael, L. Kennedy, J. L., & Mead, L. C. (1949). Some recent approaches to the experimental study of human fatigue. *Proceedings of the National Academy of Sciences, USA, 35,* 691–696.

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.

CFA (Chartered Financial Analyst) Institute. Retrieved July 20, 2008, from http://www.cfainstitute.org/cfaprogr/resources/examdetails

Chan, D., Schmitt, N., DeShon, R. P., Clause, C. S., & Delbridge, K. (1997). Reactions to cognitive ability tests: The relationships between race, test performance, face validity perceptions, and test-taking motivation. *Journal of Applied Psychology, 82,* 300–310.

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

College Board. (2007). *2007 College-Bound Seniors. Total Group Profile Report.* New York: The College Board.

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

Davis, D. R. (1946). This disorganization of behaviour in fatigue. *Journal of Neurology, Neurosurgery, and Psychiatry, 9,* 23–29.

De Vries, J., & Van Heck, G. L. (2002). Fatigue: Relationships with basic personality and temperament dimensions. *Personality and Individual Differences, 33,* 1311–1324.

Dodge, R. (1917). The laws of relative fatigue. *Psychological Review, 24,* 89–113.

D'Huyvetter, K. (1987). *Mental fatigue: The effects of time on task and mental work load on event-related potentials, subjective ratings, and task performance.* Unpublished doctoral dissertation, California State University, Long Beach.

Ebbinghaus, H. (1896–1897). 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.] (O. Wilhem, Trans.) *Zeitschrift für Psychologie und Psysiologie der Sinnesorgane* (Vol. 13, pp. 401–459).

FairTest. (2006, August). *Scores from the "New" SAT expected to dip.* FairTest: The National Center for Fair and Open Testing. Retrieved February 3, 2009, from www.fairtest.org

Faul, F., Erdfelder, E., Lang, A-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191.

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.

FOXNews.Com. (2006, May 12). *Fatigue may be reason for dip in SAT scores.* Associated Press story. Retrieved February 3, 2008, from www.foxnews.com

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.* Retrieved May 23, 2001, from http://ipip.ori.org

Grandjean, E. (1968). Fatigue: Its physiological and psychological significance. *Ergonomics, 11,* 427–436.

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

Heckhausen, H., & Kuhl, J. (1985). From wishes to action: The dead ends and short cuts on the long way to action. In M. Frese & J. Sabini (Eds.), *Goal directed behavior: The concept of action in psychology* (pp. 134–160). Hillsdale, NJ: Erlbaum.

Heggestad, E., & Kanfer, R. (2000). Individual differences in trait moti-

vation: Development of the Motivational Trait Questionnaire (MTQ). *International Journal of Educational Research, 33,* 751–776.

Hildebrand, J. (2007, August 29). A test of time. *Newsday,* p. A04.

Hockey, G. R. J. (1983). *Stress and fatigue in human performance.* New York: Wiley.

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energertical framework. *Biological Psychology, 45,* 73–83.

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

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

Huxtable, Z. L., White, M. H., & McCartor, M. A. (1946). A re-performance and re-interpretation of the Arai Experiment in mental fatigue with three subjects. *Psychological Monographs, 59,* 1–39.

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

Kaneko, K., & Sakamoto, K. (2001). Spontaneous blinks as a criterion of visual fatigue during prolonged work on visual display terminals. *Perceptual and Motor Skills, 92,* 234–250.

Kanfer, R. (1990). Motivation theory and industrial/organizational psychology. In M. D. Dunnette & L. Hough (Eds.), *Handbook of industrial and organizational psychology. Vol. 1. Theory in industrial and organizational psychology* (pp. 75–170). Palo Alto, CA: Consulting Psychologists Press.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology–Monograph, 74,* 657–690.

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 & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. 19, pp. 1–57). Greenwich, CT: JAI Press.

Kinsman, R. A., & P. C. Weiser. (1976). Subjective symptomatology during work and fatigue. In E. Simonson & P. C. Weiser (Eds.), *Psychological aspects and physiological correlates of work and fatigue* (pp. 336–405). Springfield, IL: Charles C. Thomas.

Laitusis, C. C., Morgan, D. L., Bridgeman, B., Zanna, J., & Stone, E. (2007). *Examination of fatigue effects from extended-time accommodations on the SAT Reasoning Test.* College Board Research Report No. 2007–1. New York: The College Board.

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

Lewin, K. (1935). *A dynamic theory of personality: Selected papers.* (D. K. Adams & K. E. Zener, Trans.). New York: McGraw-Hill.

Lewin, T. (2005, December 16). SAT, at 3 hours and 45 minutes, draws criticism over its length. *The New York Times,* p. 34.

Liu, J., Allspach, J. R., Feigenbaum, M., Oh, H.-J., & Burton, N. (2004). *A study of fatigue effects from the new SAT. College Board Research Report No. 2004–5.* New York: The College Board.

Lorist, M. M., Klein, M., Nieuwenhuis, S., Jong, R. D., Mulder, G., & Meijman, T. F. (2000). Mental fatigue and task control: Planning and preparation. *Psychophysiology, 37,* 614–625.

MacDonald, G. J. (2005, March 13). Early results of the new SAT: Fatigue, pressure, boredom. *USA Today*. Retrieved July 20, 2008, from http://www.usatoday.com/news/education/2005-03-13-new-sat_x.htm

Martyn, G. W. (1913). A study of mental fatigue. *British Journal of Psychology, 5,* 427–446.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S. E., Huggins, J., & Falconer, S. (1999). Validation of a comprehensive stress state questionnaire: Toward a state "Big Three?" In I. Mervielde, I. J. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 335–350). Tilburg: Tilburg University Press.

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

Mosso, A. (1906). *Fatigue.* (Margaret Drummond & W. B. Drummond, Trans.). New York: Putnam.

Muscio, B. (1921a). Feeling-tone in industry. *British Journal of Psychology, 12,* 150–162.

Muscio, B. (1921b). Is a fatigue test possible? *British Journal of Psychology, 12,* 31–46.

Myers, C. S. (1937). Conceptions of mental fatigue. *American Journal of Psychology, 50,* 296–306.

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.

Revelle, W. (1993). Individual differences in personality and motivation: "Non-cognitive" determinants of cognitive performance. In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention, selection, awareness and control: A tribute to Donald Broadbent* (pp. 346–369). Oxford: Oxford University Press.

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.

Schmeichel, B. J., Vohs, K. D., & Baumeister, R. F. (2003). Intellectual performance and ego depletion: Role of the self in logical reasoning and other information processing. *Journal of Personality and Social Psychology, 85,* 33–46.

Schmidtke, H. (1976). Disturbance of processing of information. In E. Simontke & P. C. Weiser (Eds.), *Psychological aspects and physiological correlates of work and fatigue* (pp. 336–405). Springfield, IL: Charles C Thomas.

See, J. E., Howe, S. R., Warm, J. S., & Dember, W. N. (1995). Meta-analysis of the sensitivity decrement in vigilance. *Psychological Bulletin, 117,* 230–249.

Snow, R. E. (1963). *Effects of learner characteristics in learning from instructional films.* Unpublished doctoral thesis. Purdue University. University Microfilms #6404928.

Sonnentag, S. (2003). Recovery, work engagement, and proactive behavior: A new look at the interface between work and non-work. *Journal of Applied Psychology, 88,* 518–528.

Sonnentag, S., Binnewies, G., & Mojza, E. J. (2008). "Did you have a nice evening?" A day-level study of recovery experiences, sleep, and affect. *Journal of Applied Psychology, 93,* 674–684.

Sonnentag, S., & Zijlstra, F. (2006). Job characteristics and off-job activities as predictors of need for recovery, well-being, and fatigue. *Journal of Applied Psychology, 91,* 330–350.

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

Staggs, G. D., Larson, L. M., & Borgen, F. H. (2007). Convergence of personality and interests: Meta-analysis of the Multidimensional Personality Questionnaire and the Strong Interest Inventory. *Journal of Career Assessment, 15,* 423–445.

Sullivan, B. A., & Hansen, J.-I. C. (2004). Mapping associations between interests and personality: Toward a conceptual understanding of individual differences in vocational behavior. *Journal of Counseling Psychology, 51,* 287–298.

The Economist. (2008, June 5). Charter School. Retrieved from July 20, 2008, from http://www.economist.com/finance

Thorndike, E. L. (1900). Mental fatigue. I. *Psychological Review, 7,* 547–579.

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 vali-

dation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology, 54,* 1063–1070.
Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology, 76,* 820–838.
Yeh, Y. Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors, 30,* 111–120.
Zellars, K. L., Perrewé, P. L., Hochwarter, W. A., & Anderson, K. S. (2006). The interactive effects of positive affect and conscientiousness on strain. *Journal of Occupational Health Psychology, 11,* 281–289.

Appendix

At-Home Questionnaire (AHQ)

The following scales were included in the at-home questionnaire (AHQ) packet, for the assessment of key trait complexes. 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 A1.

Table A1
Trait measures: Number of Items, Mean, SD, and Internal Consistency Reliability

Scale	No. of items	*M*	*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
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

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

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 to 15 items that were balanced in terms of positive or negative statements.

Typical Intellectual Engagement (TIE). The 59-item TIE (Goff & Ackerman, 1992) was designed to assess an individual's pattern of interests, attitudes, and behaviors regarding intellectually-demanding activities (Ackerman et al., 2001; 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 (MTQ). The short-form of the 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 (STAI). The 20-item Spielberger (1983) 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.

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 (4 to 7 items each) make up the BIS/BAS measure (BIS, BAS Reward Responsiveness, BAS Drive, and BAS Fun Seeking).

Received September 28, 2008
Revision received February 25, 2009
Accepted February 27, 2009 ■