Cognitive Fatigue During Testing: An Examination of Trait, Time-on-Task, and Strategy Influences

Phillip L. Ackerman, Ruth Kanfer, Stacey W. Shapiro, Sunni Newton & Margaret E. Beier

To cite this article: Phillip L. Ackerman, Ruth Kanfer, Stacey W. Shapiro, Sunni Newton & Margaret E. Beier (2010) Cognitive Fatigue During Testing: An Examination of Trait, Time-on-Task, and Strategy Influences, Human Performance, 23:5, 381-402, DOI: 10.1080/08959285.2010.517720

To link to this article: http://dx.doi.org/10.1080/08959285.2010.517720

Published online: 20 Nov 2010.

Article views: 321

View related articles

Citing articles: 3 View citing articles
Cognitive Fatigue During Testing: An Examination of Trait, Time-on-Task, and Strategy Influences

Phillip L. Ackerman, Ruth Kanfer, Stacey W. Shapiro, and Sunni Newton

Georgia Institute of Technology

Margaret E. Beier

Rice University

What are the consequences of testing over an extended period? We report a study of 4 hr of nearly continuous testing on two verbal tests (Cloze and Completion). Prior to the testing session, participants completed a series of nonability trait measures, including selected personality and motivation scales. During the study, participants \(N = 99\) were also administered a series of subjective fatigue and affect measures. We examined the effect of increasing time-on-task on performance and subjective fatigue, along with the relative influences of trait measures in predicting individual differences in subjective fatigue as time-on-task increased. In addition, we examined whether performance strategy differences were associated with either performance or subjective fatigue measures. Results indicated a dissociation between subjective fatigue (increasing over time-on-task) and performance measures (which were stable or showed slight improvements as time-on-task increased). Trait complexes accounted for significant amounts of variance in subjective fatigue and positive affect over the course of the test session. Performance strategies of overactivity, withdrawal, and mixed overactivity and withdrawal were identified, and correlates of the strategies were examined. Implications for analyzing performance strategies to evaluate reactions to cognitive fatigue, and the prediction of individual differences in cognitive fatigue during testing are discussed.

Two overarching questions about mental or cognitive fatigue have been raised in the context of modern applied psychology. First, can people work on a cognitively demanding test or task over an extended time without showing fatigue-related impairments? Second, are there individual differences in the propensity to experience cognitive fatigue during effortful tests/tasks over an extended period (either in terms of performance or in terms of subjective reactions)? The first question was initially addressed by Ebbinghaus (1896–97), when he created the Completion Test to be administered at various times during the school day, in an effort to determine whether the school day was too long. The second question has been raised at various times in the context of fatigue experienced during selection testing for jobs or academic placements that require a high degree of cognitive effort over long periods of time on the part of the individual (e.g., Ackerman, Kanfer, & Wolman, 2008; Whipple, 1922; Wohlhueter, 1966), but as yet there is no answer to this question.

Correspondence should be sent to Phillip L. Ackerman, School of Psychology, Georgia Institute of Technology, 654 Cherry Street, MC 0170, Atlanta, GA 30332–0170. E-mail: phillip.ackerman@psych.gatech.edu
The experience of cognitive fatigue is a phenomenon with which most, if not all, individuals have some personal familiarity. Reports of cognitive fatigue are usually associated with feelings of dullness, being weary, worn out, and so on (e.g., see Kaneko & Sakamoto, 2001). At the outset it is important to distinguish between cognitive fatigue and “boredom.” On one hand, boredom is encountered in situations where there are few stimuli that must be responded to over an extended period, sometimes in the context of high levels of stimulus “noise” or nontargets (e.g., vigilance tasks). Cognitive fatigue, on the other hand, is experienced when there are sustained demands on cognitive effort and other cognitive activities (such as thinking, problem solving, and frequent responding)—see Bartley and Chute (1947) for a discussion of these issues.

Several task characteristics have been identified as being associated with cognitive fatigue. The most prominent characteristic, and perhaps the most obvious one, is time-on-task without a break (e.g., Grandjean, 1968; Schmidtke, 1976; Thorndike, 1912). However, what constitutes a long enough period of time-on-task to result in cognitive fatigue depends on a variety of other task characteristics, such as (a) the level of demands on intellectual functioning, (b) degree of attention to detail required, (c) low tolerance for errors, (d) lack of knowledge-of-results or feedback, (e) high stakes, (f) time-pressure, and (g) a low level of intrinsic interest in the task (Ackerman & Kanfer, 2006). From this list, it is clear that there are many situations where cognitive fatigue might be encountered during testing or other performance situations, when the individual must engage in cognitively intensive activities over extended time without breaks.

**SUBJECTIVE COGNITIVE FATIGUE**

The field associated with the assessment of subjective fatigue is one that has burgeoned during the past 30 years. Research has been conducted and assessment instruments have been developed that address fatigue in the workplace (e.g., see Åhsberg, 1998; Boksem, Meijman, & Lorist, 2006) and clinical fatigue (e.g., see DeLuca, 2005; Piper et al., 1998; see also Kinsman & Weiser, 1976, for an early review). Although much of this work concerns general issues of long-term fatigue (e.g., Stein, Jacobsen, Blanchard, & Thors, 2004), there are several research programs that have investigated subjective feelings of cognitive fatigue during task performance over time. The general findings from the literature are quite clear. When the task has some or most of the characteristics just listed, and the task is administered over more than 1 or 2 hr, participants tend to report increasing levels of subjective fatigue, lower levels of positive affect, and higher levels of negative affect with increasing time-on-task (see Åhsberg, 1998). However, aside from studies of clinical populations (where the contrasts are between normals and some clinical population), there have not been any investigations to date that have examined the possibility of stable trait predictors of cognitive fatigue during a session of test performance. Thus, it is not clear whether or how personality and stable motivational traits relate to individual differences in the propensity to experience subjective fatigue during testing.

**PERFORMANCE AND COGNITIVE FATIGUE**

Historically, there have been several efforts to evaluate the performance effects associated with cognitive fatigue. Tasks have ranged from relatively simple cognitive activities, such as color naming and addition/subtraction (Bills, 1931) administered over relatively short periods (such
as a couple of hours), to highly difficult tasks, such as mental multiplication of four-digit num-
bers (e.g., Arai, 1912) over 8-hr continuous-testing periods. At the microlevel of analysis, some
studies have shown that there are brief mental “blocks” (see Bills, 1931), where the individual
slows down or lapses in attention for a brief period in the context of multiple task trials. How-
ever, these analyses and other examinations of macrolevel effects on task performance have
generally failed to find substantial decrements in performance during extended testing times.
These findings have generally led researchers to question whether there are any robust perfor-
mance effects associated with cognitive fatigue, at least within a few hours of continuous
time-on-task.

Such results appear to be discrepant with the assessments of subjective cognitive fatigue. That
is, in most evaluations of cognitive fatigue with tasks that are not intrinsically interesting (in con-
trast to tasks such as playing video games or reading for entertainment), marked negative changes
in attitudes toward the task or test, increases in self-ratings of fatigue, increases in negative affect,
decreases in positive affect, and a strong desire to discontinue the task are found, as time-on-task
increases. This dissociation between subjective fatigue and task performance represents an impor-
tant puzzle to the evaluation of cognitive fatigue. Thorndike (1900) was perhaps the first to notice
this discrepancy between increasing subjective feelings of fatigue and the lack of task perfor-
mance consequences. His interpretation was that

we can feel mentally fatigued without being so, so that the feelings described above serve as a sign
to us to stop working long before our actual ability to work has suffered any important decrease
which an experimenting psychologist could measure and use as a warning to us. (Thorndike, 1900,
p. 481)

Later, Dodge (1917) noted that these feelings of cognitive fatigue were not so much a reflection
of actual fatigue as an increasing desire to do something other than the task at hand. It is important
to note that Dodge also suggested that one can offset the feelings of fatigue by increasing one’s de-
termination and motivation to perform well on the task. Similar reports of dissociations between
subjective workload and performance have been found in other task domains (e.g., see Yeh &
Wickens, 1988).

STRATEGIES AND INDIVIDUAL DIFFERENCES

There have been few studies that have specifically targeted the assessment of individual differ-
ences in reactions to tasks that involve cognitive fatigue. The most salient of these studies was
conducted by Davis (1946), using the “Cambridge Cockpit” simulation. In this series of studies,
Davis administered an aviation task that was designed to be sufficiently difficult that adequate
performance was beyond the limits of nearly all participants. In the context of examining the
performance from a large number of participants ($N = 355$), Davis determined that there were at
least three identifiable groups, according to the respective patterns of performance. The “nor-
mal” group (approximately 75% of the sample) performed at a stable level throughout the task.
The “overactivity” group (17% of the sample) appeared to increase effort (and thus perfor-
mance) in the face of perceived difficulty in maintaining attention on the task over time. Finally,
the “withdrawal” group (the remaining 8% of the sample) showed a decrease in effort as the
time-on-task increased. These individuals also appeared to show a concomitant reduction in
commitment to the task and a downward revision of their goals. The general sense of those individuals in the withdrawal group is that they were characteristic of psychologically “leaving the field” (Lewin, 1935), a phenomenon also observed in a study by Lazarus and Eriksen (1952) in a study of testing under normal and under stressful conditions. Davis also hypothesized that some individuals in the overactivity group eventually will transition to the withdrawal strategy, when they have essentially depleted their reserves of attentional resources. Thus, a fourth potential group represents “overactivity, followed by withdrawal.” One might expect to see this pattern of results if the time-on-task is sufficient to exceed the maximal level of effort-over-time available to the participants.

The Davis investigation is specifically significant in the context of many prior and subsequent empirical failures to find overall performance decrements under fatiguing conditions. The reason for this significance is that when these different groups of individuals are combined, overall performance of the sample might be expected to remain stable, decrease, or even increase, depending on the length of time-on-task, the proportion of individuals representing each type of approach to the task, and the other task characteristics that are more or less related to the experience of cognitive fatigue.

STUDY OVERVIEW

The current investigation differs from prior research on cognitive fatigue in two ways. First, it is our goal to evaluate individual differences in cognitive fatigue in the context of cognitive test performance over extended testing time. Specifically, we examine whether distal nonability traits (personality and motivation) are related to subjective cognitive fatigue, and negative and positive affect during testing. A second goal is to determine whether the patterns of strategies described by Davis (1946) are found in a cognitive ability testing situation. In particular, we evaluate whether there are distinct patterns of performance associated with (a) a lack of fatigue effects, (b) overactivity, (c) withdrawal, and (d) a mixture of overactivity and withdrawal.

Ability Tests

The two tests selected for this study are highly similar. Based on the extant theory and literature, it should be noted that alternating between substantially different tests (such as math and verbal tests) would be expected to diminish any fatigue effects because of the variety of the tests and the differential demands on substantially different abilities. The tests we selected are the Cloze and Completion tests. Both tests provide the examinee with a text passage with every fifth word deleted and replaced with a blank. The examinee must write in the missing word, based on his or her understanding of the context of the existing text and his or her knowledge of sentence structure and grammar, and in the case of the Completion test, his or her memory of the passage. The Cloze test is generally considered to be an excellent marker for verbal fluency (e.g., see Carroll, 1993, for a review), though variants of the Cloze test are also used to assess readability of texts in domain-knowledge contexts (e.g., see Taylor, 1953). The Completion test, although a precursor of the Cloze test (Ebbinghaus, 1896–97; Terman, 1906), is essentially the same as the Cloze test, except that the full text passage is first read orally to the examinees. Thus, the Completion test represents a combination of oral comprehension and memory, in ad-
dition to the other demands of the Cloze test (because with long passages, it is not generally possible for examinees to recall a large number of the specific words). In general, a Completion test is easier than a Cloze test version of the same text passage, because hearing the entire passage provides both specific information about the missing words and information about the context of the overarching passage.

**Trait Predictors**

For the purpose of predicting subjective fatigue and negative and positive affect during testing, there are many possible nonability traits that may offer useful predictive validity. One source of individual differences that might be particularly appropriate to the study of individual differences in subjective fatigue during extended cognitive task performance over time is a broad factor of need for achievement and task mastery (which represents an approach-orientation). With a task-approach orientation, individuals are expected to be more resistant to reports of fatigue on tasks that are cognitively challenging.

Another source of individual differences is a broad factor of competitiveness/other-oriented goals. Individuals who report high levels of competitiveness and attention to other participants’ performance are expected to also report higher levels of fatigue, partly as a desire to make their own task-related efforts to appear superior to other’s efforts. These conjectures were supported in previous research on a college-student sample completing the SAT over an extended time (see Ackerman & Kanfer, 2009).

The theoretical framework underlying the investigation of these factors in areas beyond fatigue is partly based on the finding that there exists substantial communality among personality, motivation, and other traits (e.g., interests, self-concept, etc.). The framework underlying this approach was first articulated by Snow (1963). Snow introduced the concept of “aptitude complexes” as representing “combinations of levels of some variables which are particularly appropriate or inappropriate for efficient learning” (p. 120). In an examination of communality among ability, personality, and interest constructs, Ackerman and Heggestad (1997) expanded Snow’s conceptualization to include groups of traits that may be important for a wide range of criteria, not just in the domain of learning. The “trait complex” approach (see Ackerman, 1997) focuses on a relatively small number of common factors that capture the major sources of communality among personality and motivational traits. A substantial advantage of this approach is that it avoids many of the problems associated with capitalizing on chance relationships between individual trait measures and criterion measures. In addition, the trait complex approach proposes that there is a synergistic value to the combination of relatively similar traits, in terms of overarching approach or avoidance in learning and achievement contexts (e.g., see Ackerman, 2003; Ackerman & Heggestad, 1997; Armstrong, Day, McVay, & Rounds, 2008; Staggs, Larson, & Borgen, 2007; Sullivan & Hansen, 2004). The trait complexes to be derived from the measures administered in this study included those previously identified in other studies as related to ability, self-efficacy, and academic performance (e.g., Ackerman, Bowen, Beier, & Kanfer, 2001; Ackerman & Wolman, 2007), namely, trait complexes of (a) Approach-Oriented need for Achievement/Conscientiousness, and (b) Competitiveness/Attention to Others in achievement or evaluation contexts. These trait complexes have been found to be both generally coherent (in terms of shared variance) and differentially predictive of academic performance and ability criteria.
Hypotheses

The following are our hypotheses:

H1: Overall performance on the cloze and completion tests will dissociate with subjective fatigue, such that subjective fatigue will increase as time-on-task increases, while mean test performance will remain relatively stable.

H2: As subjective fatigue increases over time-on-task, it will be accompanied by increasing levels of negative affect and decreasing levels of positive affect.

H3: Competitiveness/Other-oriented goals will have a positive relationship with subjective fatigue. nAch/Conscientiousness will have a negative relationship with subjective fatigue, before, during, and subsequent to test performance.

H4: Microlevel analyses (associated with large drops or large gains in performance during testing) will find individuals who fit one of four strategy patterns: (a) Normal, (b) Overactivity, (c) Underactivity, and (d) Overactivity and Underactivity. Strategies will be common across tasks, in that strategies derived from one task will also serve as predictors of performance differences in another task given under similar circumstances.

H5: Subjective fatigue, Competitiveness/Other-oriented Goals, and nAch/Conscientiousness will be related to individual strategies during test performance over 4 hr of time-on-task. Individuals who are higher in nAch/Conscientiousness are expected to show facilitative strategies (e.g., keeping effort constant or increasing effort), although individuals higher on Competitiveness are expected to report higher levels of fatigue, which in turn relate to impeding strategies of reduced effort or an increase in effort followed by a decrease in effort over the testing session.

METHOD

Participants

Ninety-nine participants were recruited from the psychology student pool at the Georgia Institute of Technology. Inclusion criteria were normal or corrected-to-normal vision, hearing, and motor coordination, and participants were required to be native English speakers. Fifty-five men and 44 women participated, with a range in age from 18 to 24 ($M = 19.50$, $SD = 1.47$). They received academic credit for participating in the study.

Apparatus/Measures

Trait Measures

An At-Home Questionnaire packet included a variety of self-report scales designed to assess personality and motivational traits that were expected to be related to subjective fatigue and positive/negative affect during a long test session, in addition to other measures not reported here, for the sake of brevity. The scales were designed to provide robust markers for the two trait complex factors previously described. The measures are listed next.
Need for Achievement (nAch)/Conscientiousness

Eight scales were administered to assess this trait complex, as follows: (a) Conscientiousness (NEO-FFI; Costa & McCrae, 1992), (b) Need For Achievement, (c) Self-Discipline, (d) Cautiousness (International Personality Item Pool; Goldberg, 2005), (e) Mastery (Motivational Trait Questionnaire; Heggestad & Kanfer, 2000; Kanfer & Ackerman, 2000), (f) Effort Regulation, (g) Time and Study Environmental Management, and (h) Metacognitive Self-Regulation (Motivated Strategies for Learning Questionnaire; Pintrich, Smith, Garcia, & McKeachie, 1993).

Competitiveness/Other-Oriented Goals

Seven scales were administered to assess this trait complex, as follows: (a) Other-referenced Goals, (b) Competitiveness (Motivational Trait Questionnaire; Kanfer & Ackerman, 2000), (c) Extrinsic Goal Orientation (Motivated Strategies for Learning Questionnaire; Pintrich et al., 1993), (d) Agreeableness (Reversed), NEO-FFI (Costa & McCrae, 1992), (e) Self-Consciousness (Fenigstein, Scheier, & Buss, 1975), (f) Behavioral Activation System – Drive, and (g) Behavioral Activation System–Fun Seeking (BIS/BAS; Carver & White, 1994).

Composite scores were formed for each trait complex by converting each scale score to a $z$ score and then summing the scores for each scale within a trait complex. Table 1 reports means, standard deviations, and coefficient alphas for each scale, and coefficient alphas for the scales that make up the trait complexes.

<table>
<thead>
<tr>
<th>Scale</th>
<th>No. of Items</th>
<th>$M$</th>
<th>$SD$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scales for nAch/Conscientiousness$^a$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Conscientiousness (NEO-FFI)</td>
<td>12</td>
<td>50.27</td>
<td>8.55</td>
<td>.86</td>
</tr>
<tr>
<td>2. Need for Achievement (IPIP)</td>
<td>10</td>
<td>45.69</td>
<td>7.06</td>
<td>.87</td>
</tr>
<tr>
<td>3. Self-Discipline (IPIP)</td>
<td>9</td>
<td>30.44</td>
<td>7.36</td>
<td>.87</td>
</tr>
<tr>
<td>4. Cautiousness (IPIP)</td>
<td>7</td>
<td>28.32</td>
<td>5.22</td>
<td>.82</td>
</tr>
<tr>
<td>5. Mastery (MTQ)</td>
<td>8</td>
<td>35.22</td>
<td>5.70</td>
<td>.84</td>
</tr>
<tr>
<td>6. Effort Regulation (MSLQ)</td>
<td>4</td>
<td>16.45</td>
<td>3.70</td>
<td>.76</td>
</tr>
<tr>
<td>7. Time and Study Environmental Management (MSLQ)</td>
<td>8</td>
<td>31.42</td>
<td>6.47</td>
<td>.77</td>
</tr>
<tr>
<td>8. Metacognitive Self-Regulation (MSLQ)</td>
<td>12</td>
<td>45.61</td>
<td>8.07</td>
<td>.79</td>
</tr>
<tr>
<td>Scales for Competitiveness/Other-oriented goals$^b$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Other-oriented goals (MTQ)</td>
<td>7</td>
<td>30.66</td>
<td>5.60</td>
<td>.85</td>
</tr>
<tr>
<td>2. Competitive Excellence (MTQ)</td>
<td>6</td>
<td>23.50</td>
<td>6.32</td>
<td>.90</td>
</tr>
<tr>
<td>3. Extrinsic Goals (MSLQ)</td>
<td>4</td>
<td>18.91</td>
<td>3.76</td>
<td>.78</td>
</tr>
<tr>
<td>4. Agreeableness (NEO-FFI)</td>
<td>12</td>
<td>49.34</td>
<td>8.25</td>
<td>.82</td>
</tr>
<tr>
<td>5. Self-Consciousness$^c$</td>
<td>5</td>
<td>21.57</td>
<td>3.93</td>
<td>.74</td>
</tr>
<tr>
<td>6. BAS Drive</td>
<td>4</td>
<td>3.99</td>
<td>.98</td>
<td>.85</td>
</tr>
<tr>
<td>7. BAS Fun Seeking</td>
<td>4</td>
<td>4.24</td>
<td>.95</td>
<td>.84</td>
</tr>
</tbody>
</table>

Note. $\alpha$ = internal consistency reliability. NEO-FFI = NEO Five-Factor Inventory; IPIP = International Personality Item Pool; MTQ = Motivational Trait Questionnaire; MSLQ = Motivated Strategies for Learning Questionnaire; BAS = Behavioral Activation System.

$^a$Number of scales = 8, trait complex $\alpha = .91$. $^b$Number of scales = 7, trait complex $\alpha = .76$. $^c$Fenigstein et al. (1975).
Pretest, Interim, and Posttest Questionnaires

A questionnaire of fatigue and state affect (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 five times during the laboratory session. The measure included a subjective fatigue scale that included four statements (e.g., “I am having difficulty keeping my eyes open”), with responses on a 6-point Likert-type scale, from 1 (strongly disagree) to 6 (strongly agree). Scales of positive affect (6 items; e.g., “Energetic,” “Enthusiastic”) and negative affect (12 items; e.g., “Uneasy,” “Upset”), with responses on a 5-point Likert-type scale, from 1 (very slightly or not at all) to 5 (extremely), were administered as the same time as the fatigue scale.

To evaluate the subjective representation of effort management strategies along the lines of the strategies mentioned by Davis (1946), at the end of the session, participants were asked about their strategies in managing their effort during the session. The participants were asked whether they “increased my effort,” “kept my effort at a constant level,” “decreased my effort,” or “first I increased my effort, then I later decreased my effort.”

Other scales were administered in this questionnaire but are not reported here, for the sake of brevity.

Ability Tests

The two tests administered in this study, Cloze and Completion, are generally considered to assess verbal abilities in general, and verbal fluency in particular. In addition, there is an immediate memory component to the Completion Test (e.g., see Ackerman, Beier, & Bowen, 2000). Passages used to create the Cloze and Completion tests were selected from SAT preparation books, obtained from the following sources: College Board SAT, 2004; Kaplan Test Prep and Admissions, 2004; The Princeton Review, 2005. Following the technique originated by Taylor (1953), “structural” (Ohnmacht, Weaver, & Kohler, 1970) Cloze tests were constructed. This entailed leaving the first and the last sentences of the passage intact. Starting with the second sentence, every fifth word was deleted (regardless of its grammatical or contextual relationship) and replaced with an underlined blank 10 spaces long. A set of 68 passages was prepared, containing between 36 and 48 blanks per passage. Passages were then subjected to pilot testing and organized according to the level of difficulty based on the mean score for each passage. In preparing the final set of passages, efforts were made to select sets of passages that were matched in terms of number of blanks and level of difficulty. The tests were then matched into sets in terms of average number of blanks and mean score, and then randomly selected to be administered as Cloze or Completion formats.

Cloze Tests

For the Cloze tests, participants were instructed to read through the passage and fill in the blanks with the words that best fit into the sentence. If participants did not know the exact words that fit in the blank, they were instructed to guess. Participants were given 6 min to complete each test. Credit was given an exact word match (2 points) or for words that fit the gist of the paragraph
and were grammatically correct in the context of the text (1 point). Sixteen tests were administered over the course of the testing session.

**Completion Tests**

Completion tests differed from Cloze tests only in their administration (based on a design introduced by Terman, 1906). Specifically, participants were instructed to listen to the passage read in its entirety, without looking at the Completion test form. After the passage was read, participants were shown the Completion test form and instructed to fill in as many of the missing words as possible. If they did not remember the exact words, participants were instructed to guess. Participants were given 6 min to fill in the Completion test form. Scoring for the Completion tests was identical to that of the Cloze tests. Sixteen tests were administered over the course of the testing session.

Order of the tests across the four 1-hr sets was counterbalanced across participants with four different orders selected from a Latin Square table.

**Procedure**

After signing up for the study, participants received the at-home questionnaire via mail, up to 1 week prior to the laboratory session. All participants were required to complete the questionnaire and bring it with them to the laboratory session. All laboratory sessions started at 9:00 a.m. and ran for 4½ hr. Participants first received a general overview of the tests, and completed the first interim questionnaire, prior to completing any of the tests. The questionnaire was followed by an alternating set of four Cloze and four Completion tests (which lasted 56 min), the second interim questionnaire, a second alternating set of four Cloze and four Completion tests, followed by a 5-min break. On return from the break, participants completed the third interim questionnaire, followed by another set of Cloze and Completion tests, the fourth interim questionnaire, the fourth set of Cloze and Completion tests, and finally a posttest questionnaire. Participants received no feedback or knowledge of results on these tests throughout the session. Participants were then debriefed and excused.

**RESULTS**

The results are presented in five sections. In the first section, we report changes in performance and self-reports of fatigue and affect over the course of the testing session. In the second section, we review correlations between self-report measures and test performance measures. In the third section, we consider the relations between trait complexes (including personality and motivation constructs) and self-reported fatigue and affect during testing. In the fourth section, we evaluate the effects of time-on-test on test performance, in terms of within-hour analyses and specific performance strategies. In the final section, we report a model of subjective fatigue across the session, including trait-complex determinants of subjective fatigue. For reference purposes, we also report intercorrelations among the predictor and criterion variables in Table 2.
TABLE 2  
Correlations Among Predictor and Criterion Measures

|       | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  | 14  | 15  | 16  | 17  | 18  | 19  | 20  | 21  |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1. nAch/Mastery |     | .053 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 2. Competitiveness | .262 | .271 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 3. Subjective Fatigue1 | -.262 | .271 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 4. Subjective Fatigue2 | -.180 | .323 | .773 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 5. Subjective Fatigue3 | -.189 | .296 | .572 | .785 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 6. Subjective Fatigue4 | -.126 | .292 | .464 | .675 | .811 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 7. Positive Affect1 | .173 | -.070 | -.386 | -.356 | -.244 | -.248 |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 8. Positive Affect2 | .200 | -.172 | -.412 | -.576 | -.477 | -.388 | .771 |     |     |     |     |     |     |     |     |     |     |     |     |
| 9. Positive Affect3 | .206 | -.162 | -.306 | -.405 | -.530 | -.494 | .596 | .714 |     |     |     |     |     |     |     |     |     |     |     |
| 10. Positive Affect4 | .270 | -.173 | -.196 | -.312 | -.362 | -.478 | .498 | .605 | .713 |     |     |     |     |     |     |     |     |     |     |
| 11. Negative Affect1 | -.197 | .086 | .261 | .050 | .096 | .153 | -.217 | -.129 | -.130 | -.081 |     |     |     |     |     |     |     |     |     |
| 12. Negative Affect2 | -.206 | .165 | .350 | .299 | .346 | .383 | -.188 | -.263 | -.144 | -.112 | .741 |     |     |     |     |     |     |     |     |
| 13. Negative Affect3 | -.159 | .116 | .287 | .186 | .300 | .382 | -.249 | -.241 | -.168 | .689 | .836 |     |     |     |     |     |     |     |
| 14. Negative Affect4 | -.082 | .112 | .247 | .178 | .261 | .432 | -.264 | -.205 | -.194 | .215 | .583 | .739 | .895 |     |     |     |     |     |     |
| 15. Completion1 | -.075 | -.320 | -.189 | -.266 | -.143 | -.160 | .030 | .088 | .077 | .048 | -.015 | -.150 | -.068 | -.086 |     |     |     |     |     |
| 16. Completion2 | -.027 | -.285 | -.268 | -.433 | -.353 | -.372 | .002 | .143 | .135 | .065 | .016 | -.136 | -.073 | -.097 | .771 |     |     |     |     |
| 17. Completion3 | .000 | -.292 | -.273 | -.383 | -.316 | -.296 | -.007 | .133 | .150 | .110 | .019 | -.101 | -.037 | -.043 | .720 | .892 |     |     |     |
| 18. Completion4 | .040 | -.307 | -.240 | -.308 | -.267 | -.271 | .046 | .146 | .154 | .121 | .024 | -.165 | -.098 | -.149 | .733 | .754 | .808 |     |     |
| 19. Cloze1 | -.030 | -.283 | -.221 | -.260 | -.178 | -.165 | -.001 | -.006 | .057 | .038 | .044 | -.043 | -.040 | -.057 | .736 | .747 | .723 | .645 |     |
| 20. Cloze2 | .045 | -.270 | -.164 | -.243 | -.165 | -.146 | .000 | .108 | .072 | .055 | -.016 | -.171 | -.042 | -.034 | .627 | .730 | .769 | .722 | .635 |
| 21. Cloze3 | -.062 | -.248 | -.101 | -.126 | -.175 | -.183 | -.027 | .110 | .085 | .045 | -.003 | -.164 | -.045 | -.066 | .659 | .571 | .601 | .690 | .429 | .607 |
| 22. Cloze4 | -.058 | -.215 | -.254 | -.228 | -.170 | -.187 | -.015 | .035 | .066 | -.007 | .000 | -.104 | -.087 | -.090 | .758 | .703 | .693 | .669 | .681 | .514 | .708 |

*Note.*  
N = 99. Correlations larger than ± .197, significant p < .05. nAch = Need for Achievement.
Mean Effects of Time-on-Task on Performance

Completion test performance showed negligible overall performance changes over time-on-task, $F(3, 291) = 1.65, ns$, comparison between first and last hour of performance (Cohen’s $d = .02$), and Cloze test showed an overall small improvement, $F(3, 291) = 4.06, p < .01$, comparison between first and last hour of performance ($d = .04$). Together, time-on-task effects accounted for less than 5% of the variance in performance across the Cloze and Completion tests. See Figure 1 for means over each hour of testing.

Mean Effects of Time-on-Task on Self-Reports

The self-report measures administered during the testing session provided broad agreement with the general expectation that performing Cloze and Completion tests nearly continuously over a period of 4 hr was not a particularly pleasant experience—see Table 3. Subjective fatigue showed a significant increase from the first measurement (prior to the 1st hour of testing) to the fourth measurement (prior to the 4th hour of testing). In addition, significant decreases in positive affect were found, along with significant increases in negative affect during the same period. Subjective fatigue showed an overall increase over time-on-task (The effect size for time-on-task was $\eta^2 = \ldots$)

![FIGURE 1 Mean performance on the Completion Tests and Cloze Tests (averaged over four tests/hour) over the course of 4 hr of testing.](image-url)
and there were concomitant decreases in positive affect and increases in negative affect (time-on-task accounted for 26% and 10% of the respective score variances). For the self-report measures administered after all testing was completed, there was a significant increase in subjective fatigue and positive affect (in comparison to the last measurement during testing) and a significant decrease in negative affect (see Table 3 for details), consistent with an interpretation that the participants were fatigued but quite relieved to be “done” with the experiment.

Examination of the performance and self-report measures together, indicated a clear dissociation, as predicted in H1. Subjective fatigue increased, negative affect increased, and positive affect decreased with increasing time-on-task, yet mean test performance remained stable (in the case of the Completion test) or slightly increased (in the case of Cloze test).

Correlations Among Self-Report Measures

Subjective fatigue, positive affect, and negative affect measures tracked well—that is, as subjective fatigue increased with time-on-task, positive affect decreased and negative affect increased (see Table 3), consistent with H2. It is also useful to examine the relations among these variables. The synchronous (i.e., measured at the same time-on-task) correlations between subjective fatigue and positive affect were negative and significant ($r = -.38$, $-.58$, $-.53$, and $-.48$ for Hours 1–4, respectively). For subjective fatigue and negative affect, the correlations were positive and significant, though of a somewhat lower magnitude ($r = .26$, $.30$, $.30$, and $.43$), all of which were significant at $p < .01$. Correlations among positive and negative affect measures were significant, but even smaller in magnitude ($r = -.22$, $-.26$, $-.24$, and $-.21$, respectively, all $p < .05$). These results suggest that subjective fatigue measures are providing unique information beyond positive affect and negative affect but that there is significant overlap among these three subjective measures.

Correlations Between Trait Complexes and Self-Report Measures

The trait complexes provide a window into the relations between distal personality and motivational traits and proximal feelings during the extended Cloze and Completion testing. The raw

### Table 3

<table>
<thead>
<tr>
<th>No. of Items</th>
<th>M (SD)</th>
<th>α</th>
<th>1st hr</th>
<th>2nd hr</th>
<th>3rd hr</th>
<th>4th hr</th>
<th>F</th>
<th>MSE</th>
<th>ES</th>
<th>Posttest</th>
<th>t</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Fatigue</td>
<td>4</td>
<td>.81</td>
<td>2.68</td>
<td>3.18</td>
<td>3.13</td>
<td>3.29</td>
<td>19.80**</td>
<td>.35</td>
<td>.17</td>
<td>3.47</td>
<td>−2.71**</td>
<td>.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.94)</td>
<td>(1.06)</td>
<td>(1.11)</td>
<td>(1.10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Positive affect</td>
<td>6</td>
<td>.76</td>
<td>2.23</td>
<td>1.90</td>
<td>1.88</td>
<td>1.70</td>
<td>35.20**</td>
<td>.14</td>
<td>.26</td>
<td>2.11</td>
<td>−5.39**</td>
<td>.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.59)</td>
<td>(.56)</td>
<td>(.68)</td>
<td>(.64)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Negative affect</td>
<td>12</td>
<td>.90</td>
<td>1.47</td>
<td>1.58</td>
<td>1.63</td>
<td>1.74</td>
<td>10.48**</td>
<td>.12</td>
<td>.10</td>
<td>1.66</td>
<td>2.40*</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.55)</td>
<td>(.63)</td>
<td>(.75)</td>
<td>(.79)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Number of items and internal consistency for the first administration, along with means, standard deviations, and analysis of variance (ANOVA) test results. Also provided are $t$ tests between measures taken before the 4th hr and after the last tests were completed. $F (3, 294) =$ repeated measures ANOVA for four interim questionnaires; $t =$ dependent $t$ test comparing final (4th hr) interim questionnaire with posttest questionnaire ($df = 98$); ES = partial eta squared; $d =$ Cohen’s $d$. 

...
correlations are provided in Table 2. In addition, multiple correlations were computed for predicting subjective fatigue and positive/negative affect from the trait complex measures. The trait complexes provided significant predictive validity for subjective fatigue before, during, and after the test session \((R^2 = .15, .14, .13, .10, \text{and} .10\) for Hours 1–4 and the posttest assessment, respectively, all \(p < .01\)). The prediction of positive affect from the trait complexes was not significant for the first hour but was significant in subsequent assessments \((R^2 = .04, .07, .07, .11, \text{and} .13, \) respectively, all \(p < .05\)). The prediction of negative affect was only significant at Hour 2 \((R^2 = .05, .07, .04, .02, \text{and} .05, \) respectively). The trait complex prediction of positive affect was not significantly greater than the prediction of negative affect, given the modest correlations between the two measures. Together, these results provide partial support H3. That is, trait-complex measures taken several days before the actual testing situation were significantly predictive of subjective fatigue and positive affect but not for negative affect immediately before, during, and after testing in a 4-hr test session.

**Strategies and Effort Management**

Analysis of the strategy question responses at the end of the session indicated that half of the participants reported keeping their effort at a constant level (51), and fewer participants reported increasing their effort (15), decreasing their effort (10), or increasing, then decreasing effort (22). These different strategies were significantly related to reported levels of subjective fatigue across the session, \(F(3, 94) = 2.73, p < .05, \text{MSE} = 38.64\). The highest levels of fatigue were reported by the participants who indicated that they increased in effort, and the lowest levels of fatigue were reported by participants who reported keeping a constant level of effort across the session.

**Within-Session Analyses**

Although the hourly averages of performance in the Cloze and Completion tests reflect relatively consistent performance, examination of performance on the individual tests within each hour provides a quite different impression. Within each hour, there was a drop in performance from the second to the fourth test, which became more pronounced after the first hour, as shown in Figure 2. The effects of position within hour was significant, \(F(9, 882) = 10.99, p < .01, \eta^2 = .10, \text{and} 33.68, p < .01; \eta^2 = .26, \) for the Completion and Cloze tests, respectively. Mean performance on the individual tests (see Figure 2) shows that, especially for the Cloze test, the first two tests within the hour (after the first hour) showed markedly higher levels of performance than the last two tests within the hour. Performance on the Completion test showed a somewhat different pattern, with performance dropping from the first to the third test (after the 1st hour), and the last test in each hour indicating some marginal improvements.

The question to be answered from the within-session analysis is whether there are marked individual differences in the tendency for performance to substantially drop off as each hour proceeds in testing. For example, if most or all of the participants have the same pattern of declining performance during each hour, then there is a general effect of fatigue, but not something that likely would be associated with either the trait complexes or the self-reports of fatigue, or with positive or negative affect. In contrast, if there are individual differences in how people responded to the extended testing session, they may be related to these self-report measures and trait complexes. If
so, it might be possible to use these measures to anticipate which individuals are going to be more or less susceptible to fatigue effects.

The earlier fatigue literature suggests that there are perhaps four types of typical responses or strategies used in the context of performance over extended periods (see Davis, 1946), as follows:

1. No change. The first response is that which would be associated with an individual who does not feel fatigued. Such an individual could be expected to keep his or her performance relatively level across testing sessions.

2. Perceived fatigue followed by decreasing effort. In this case, the individual feels fatigued and reduces his or her effort to best preserve limited cognitive resources. Performance under this response type would be expected to decline with increasing time-on-test.

3. Perceived fatigue followed by increasing effort. In this case, the individual has feelings of fatigue, but the reaction is a positive one (increasing effort to maintain a perceived level performance) instead of withdrawing effort. In this case, performance is expected to either remain level or actually increase, depending on the amount of extra effort put forth by the individual and the underlying performance-effort function (e.g., see Kanfer, 1987).

4. Mixed. Finally, there is a mixed pattern that underlies a perception of fatigue, and an initial increase in effort to counteract fatigue, followed by a drop-off in effort as the individual reaches a

---

**FIGURE 2** Mean performance on the individual Completion Tests and Cloze Tests over the course of 4 hr of testing.
point where continued effort becomes overly aversive. In this case, performance might stay level and then drop off, or might show an increase followed by a steep decrease. It is also important to note that the time course of perceived fatigue and reactions to it may differ between individuals, both generally (in terms of tolerance for mental effort) and in a particular situation (depending on the individual’s initial sense of fatigue prior the beginning of the study). Thus, these patterns might be seen in any or all of the four hours of testing.

To examine these issues, we focused on the Cloze test (which appeared to be most sensitive to fatigue effects). Three measures were derived from the individual test performance data. The first measure, “Drop,” was an indication of whether an individual showed a drop of more than 20 percentage points in test performance from the second to the fourth Cloze test in each hour. (This represents a drop in performance greater than 1 SD unit in all cases.) A breakdown of participants on this variable indicated that 75 of the participants did not show a substantial drop in performance during any of the 4 hr, whereas 24 of the participants did. Five participants showed a Drop in the 1st hr of testing, whereas 8, 4, and 8 showed Drops in Hours 2, 3, and 4, respectively. (The numbers add to 25 because 1 participant showed a Drop in both the 2nd and 4th hr).

We also calculated a “Spurt” measure, which was based on whether an individual showed an increase in performance during the same periods of more than 20 percentage points. For this measure, 41 participants did not show a Spurt in performance, whereas 58 of the individuals did show such a Spurt at some point in the test hours. The distribution of Spurts across the hours showed 11 in the 1st hour, 17 in the 2nd hour, 18 in the 3rd hour, and 15 in the 4th hour. Three individuals showed more than one Spurt in performance across the testing hours.

A cross-tabulation of the Drop × Spurt variables indicated 31 participants with no Drop and no Spurt, 44 participants with no Drop but a Spurt in performance, 10 participants with a Drop in performance but no Spurt, and 14 participants with a Drop and a Spurt in performance. There was no significant relationship between these Drop and Spurt variables (Cramer’s V = .003, ns). An examination of the overall test performance (shown in Figure 3) of these groups (which were formed solely on the basis of Cloze test Drops and Spurts) illustrates significant two-way interactions between Drops and Spurts for the Cloze test, \( F(1, 95) = 5.51, p \lt .05, \eta^2 = .06, \text{MSE} = 402.88 \). It is important to note, however, that these results might be at least partially due to regression-to-the-mean effects. Thus, we also compared the performance of these different groups, formed from Cloze test scores, to the scores on the Completion tests. We also found a similar interaction among Drops and Spurts, for performance on the Completion test, \( F(1, 95) = 5.13, p < .05, \eta^2 = .05, \text{MSE} = 500.10 \). Although participants who showed relatively consistent performance performed well in both tests, those participants who showed a substantial Drop during the testing sessions, but who also made up performance by showing Spurts of recovered performance, did quite well, too. The participants who showed a substantial Drop during test performance without any notable Spurt performed the worst overall, with the individuals who showed a Spurt but not the substantial Drop performed at an intermediate level. It is important to note that these group differences were not apparent or significant at the first testing occasion for either test. These results support H4, that is, groups of individuals were found who appeared to have one of the four different strategies during testing.

No significant differences were found between the Drop and No Drop groups on the final session effort management strategy question. However, an analysis of Spurt and No Spurt groups indicated a significant effect of strategy (\( \Phi = .29, p < .05 \)). Participants who reported that they kept
their effort at a constant level were more than twice as likely to show spurts in performance, compared to the other participants.

Although the power of further tests of these variables is relatively low (given the small numbers of participants in each cell), and the robustness of the results is also low (given the unequal cell sizes), the Competitiveness/Other-Oriented Goals trait complex was related to the Spurt measure ($r$ [point biserial] = .22, $p < .05$)—individuals who were higher on Competitiveness were more likely to show a Spurt in performance. None of the self-report measures administered during the testing session had significant relations with the Drop and Spurt measures. The results also provided some partial exploratory support for H5, that is, that the subjective fatigue and Competitiveness trait complex provide significant associations with strategy differences during testing.

**Analysis of Change in Subjective Fatigue**

Latent Growth Curve Modeling allows the examination of variability in initial performance and growth trajectories on variables measured within person over time (Bollen & Curran, 2006; Singer
This method also permits an examination of individual differences variables that might account for within-person variance in initial performance and growth. MPlus version 4.2 (Muthén & Muthén, 2006) was used to examine growth trajectories for fatigue (unconditional model), and individual differences in traits complexes that account for variance in initial levels of fatigue and changes in fatigue over time (conditional models; Bollen & Curran, 2006).

We first analyzed the unconditional growth model to establish the trajectory of growth for self-reported fatigue over the course of the 4-hr study. Examination of the means and the individual trajectories suggested a quadratic growth pattern, and we tested our unconditional model with three latent factors representing intercept, linear growth, and quadratic growth, respectively. In this model, path coefficients for all intercepts were set to 1; coefficients for linear slope were set to 0, 1, 2, 3, and 4; and coefficients for quadratic slope were set to 0, 1, 4, 9, and 16. Latent factors for intercept, slope and quadratic trends were allowed to covary. Error variances for each of the five time points were set to be equal (i.e., we assumed that errors were distributed independently and homoscedastically over time within person). We also assumed that prior levels of fatigue would influence future levels of fatigue (autoregression; Zyphur, Chaturvedi, & Arvey, 2008). Fit of this model was good, $\chi^2(6, N = 99) = 6.57, p > .05$, comparative fit index (CFI) = .99, root mean square error of approximation (RMSEA) = .031. The mean intercept was significant (i.e., $M = 2.68, SE = .094$), but mean values were not significant for slope or quadratic factors ($M_{\text{slope}} = –.54, SE = .41; M_{\text{quad}} = .15, SE = .096$). Although the nonsignificant mean values suggest that there was not systematic growth in fatigue over time, there was significant random variance in intercept ($\sigma^2_{\pi_0} = .632, SE = .13$), slope ($\sigma^2_{\pi_1} = .26, SE = .089$) and quadratic factors ($\sigma^2_{\pi_2} = .01, SE = .005$). As such, our next step was to examine the individual differences factors that might account for this variance.

The conditional model was identical to the unconditional model just described but included the trait complexes as exogenous predictors of variability in intercept, slope, and quadratic factors. This model is shown in Figure 4. Fit of the model was good, $\chi^2(10, N = 99) = 8.62, p > .05$, CFI = 1.0, RMSEA = .00. As can be seen in the figure, both the Need for Achievement and Competitiveness trait complexes accounted for significant random variance in initial levels of fatigue, but neither trait complex accounted for changes in fatigue over time. We also found that people who reported less fatigue initially had a greater increase in fatigue throughout the study (i.e., initial levels of fatigue were negatively correlated with the growth of fatigue). Also, those experiencing the most growth in fatigue were significantly more likely to report that their fatigue leveled off over the course of testing (i.e., slope and quadratic factors were significantly negatively correlated).

**DISCUSSION AND CONCLUSIONS**

Over the past few decades, surprisingly little research attention has been given to the relationship between cognitive fatigue and test performance. In practice, however, the topic has become increasingly important as students spend more classroom time on testing, and as some high-stakes testing applications (e.g., the SAT) require longer testing sessions (e.g., see Ackerman & Kanfer, 2009). Predicting who may experience cognitive fatigue and the relationships among subjective fatigue, behavior, and test performance have potentially important implications for the design of multiaptitude and achievement test batteries, how students who feel
fatigued during testing are identified, and how accommodations for subjective fatigue might be constructed. In this study we sought to reinvigorate and extend research in this area by integrating older research findings with contemporary approaches and measures. Our findings help to reconcile some previous empirical inconsistencies and suggest several promising new directions for future theory and research.

Specifically, our results provide evidence to support the utility of investigations that take non-ability traits into account and examine the strategies by which individuals respond to cognitively demanding tests over time. Consistent with previous research, we found a disassociation between mean level reports of subjective cognitive fatigue and test performance. That is, although subjective cognitive fatigue increased with time-on-test, global test performance remained relatively stable over the lengthy performance period. Examination of time-on-test effects from an individual difference perspective, however, revealed significant relations between certain non-ability traits and the subjective experience of fatigue. These traits, in turn, were significantly related to distinctive strategies that individuals used to achieve performance within each hour of performance. Taken together, the findings help to explain observed mean level effects, to clarify who may be expected to experience higher levels of subjective cognitive fatigue over time-on-test, and the multiple pathways by which subjective fatigue influences behavior and, in turn, test performance.

FIGURE 4 Latent growth curve model of subjective fatigue scores with intercept, slope, and quadratic factors predicted by individual differences in trait complexes.
The initial results for the Cloze and Completion tests appear, at least on the surface, to indicate that there are no detrimental effects on test performance of 4 hr of nearly continuous testing on verbal fluency and memory tests. Indeed, there was a small but significant improvement in mean performance on the Cloze tests across the 4 hr of testing.

Subjective Reactions and Subjective Fatigue

Subjective reports of the participants point to a radically different conclusion from that derived from 4-hr averages of objective test performance. Over the course of 4 hr of testing, the participants reported significantly and substantially higher levels of subjective fatigue, higher levels of negative affect, and lower levels of positive affect. The larger reduction of positive affect, compared to the magnitude of increase in negative affect, might indicate that fatigue has a more pronounced effect on reducing positive emotions that in does in increasing negative emotions, but this issue requires more study. In comparison to changes in subjective fatigue, the affect reactions were also robust, and they only dissipated (to some degree) at the end of the experiment—when the participants no longer had to complete any additional tests.

One main goal of this investigation was to determine whether there were any distal trait predictors of self-reported fatigue or other reactions to testing over an extended period. Based on the trait-complex framework, a relatively clear pattern of results was obtained. Trait complex composites together accounted for 10 to 15% of the variance in subjective fatigue before, during, and after the 4½-hr session. (This can be contrasted with the effect size for changes in subjective fatigue during testing, where time-on-task accounted for 17% of the variance in subjective fatigue.) That is, measures administered prior to testing accounted for roughly the same amount of variance in subjective fatigue as the experimental manipulation of 4 hr of nearly continuous time-on-task.

Positive affect during testing were also predicted to a significant degree by other personality and motivational trait complexes (4 to 11% of the variance in positive affect). From these results, the picture that emerges is one of overarching associations between general personality and motivational traits and the individual’s tendency toward positive or negative experiences in a testing situation.

Although the sample of participants for this study was from a population with a restriction-of-range in ability, the results are suggestive for implications regarding test accommodations that involve extended testing times. From these results, we would speculate that individuals who have high levels of general anxiety/test anxiety may find difficulty with situations that allow for extra test completion time in a single test session. That is, higher levels of anxiety appear to be related to higher levels of subjective fatigue during testing. If high-anxiety individuals are given extra time to complete a test, they may, in turn, have even higher levels of subjective fatigue than individuals who complete the test with normal time limits. Thus, the extra time allocated may not be as beneficial to those individuals as, say, dividing the test into parts across two or more separate sessions.

Performance Strategies

The identification of participants showing the normal (i.e., unaffected by the fatiguing conditions), overactivity (increasing performance), withdrawal (decreasing performance), and a mixture of overactivity and withdrawal performance strategies provided some additional insight into
the underlying stable performance of the overall group. Although 30% of the participants showed no indication of overactivity or withdrawal, 44% of the participants indicated at least one period of overactivity with no indication of withdrawal, 10% of the participants showed withdrawal in performance, and 14% of the participants appeared to show both overactivity and withdrawal. The best performance across the two tests was obtained by the “normal” group, but the next best performance was obtained by the participants who had both periods of overactivity and withdrawal. The worst performance was obtained by those individuals who showed only withdrawal in the course of testing.

The current study does not provide definitive evidence that it is possible to predict which examinees in an extended testing session are likely to experience greater subjective fatigue or show depressed test performance. However, the patterns of correlations between the trait complex measures, subjective fatigue, and subjective reactions during testing provide an indication that individual differences in distal nonability traits may be partly responsible for some negative reactions to testing over an extended period. In addition, by identifying different performance strategies along the lines suggested by Davis (1946) in his study of pilots, it is possible to see performance effects associated with extended testing and to start the process of identifying which individuals may be more or less likely to respond to such a situation with overactivity or withdrawal behaviors, either of which may have much broader implications for selection and for test validity. The finding that individual differences in Competitiveness were related to these different strategies suggests further that the role of personality and motivational traits may be more likely to be found in test-taking strategies than with an absolute level of performance. The implications for situations when cognitive ability testing is extended over a long time are less clear. The context of the current study was not a high-stakes testing situation, and participants might be expected to be more predisposed to withdrawal behaviors when the test results are not important to their academic goals (although this consideration would not account for the overactivity strategies exhibited by many participants). In addition, it is not obvious whether part of the validity of such tests for academic and occupational selection is, as suggested by Thorndike, partly a matter of ability and partly a matter of perseverance (see Whipple, 1922). Nonetheless, we think that this approach to examining cognitive fatigue has proved to be exceptionally promising for future examination of these issues.

In the final analysis, there are two take-home messages from this work. First, personality and motivational trait factors account for a significant portion of the variance in subjective fatigue, before, during and after task practice on an extended testing session. Taking account of predictors of individual differences in subjective fatigue in future studies may allow investigators to better separate task-induced fatigue effects from trait factors that may indicate how participants assess their reactions to the testing situation. For example, individuals who are more likely to anticipate greater levels of subjective fatigue may adopt different effort strategies, which in turn affect the overall outcome of test performance, especially when the test session lasts several hours or longer.

The second point is that although high-level analysis of cognitive task performance (e.g., hourly averages) may not indicate effects associated with fatigue, patterns of performance on shorter time scales (e.g., tests within hourly sessions) can illuminate fatigue effects, especially as they relate to individual strategies and patterns of effort regulation. Choosing the appropriate level of analysis for examining fatigue effects may ultimately allow for a better understanding of the historical puzzle of the dissociation between how individuals perceive the testing experience and how they actually perform on the tasks in question.
ACKNOWLEDGMENTS

Stacey Shapiro is now at Russell Reynolds Associates, Atlanta, Georgia.

REFERENCES


