REDUCING ADVERSE IMPACT: AN INVESTIGATION OF THE EFFECT OF ADDITIONAL STUDY TIME ON TRAINABILITY TEST PERFORMANCE

A Dissertation

Presented to

The Graduate Faculty of The University of Akron

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

James E. Kuthy

December, 2008
REDUCING ADVERSE IMPACT: AN INVESTIGATION OF THE EFFECT OF ADDITIONAL STUDY TIME ON TRAINABILITY TEST PERFORMANCE

James E. Kuthy

Dissertation

Approved:
Advisor
Dennis Doverspike

Accepted:
Department Chair
Paul E. Levy

Committee Member
Paul E. Levy

Dean of the College
Ronald F. Levant

Committee Member
Andrea F. Snell

Dean of the Graduate School
George R. Newkome

Committee Member
Harvey L. Sterns

Date

Committee Member
James R. Rogers

ii
ABSTRACT

The current study investigated characteristics and behaviors believed to influence performance on a Test Preparation Manual (TPM) style test used for the selection of firefighters in order to determine whether the TPM resulted in lower adverse impact when compared to a measure of cognitive ability, which was administered during the same session. TPM-style testing is a type of trainability testing process during which job applicants study a job-related manual over an extended period of time and then answer questions from memory about information contained in that manual. The characteristics and behaviors investigated included the amount of time participants reported studying a firefighter training manual on which they were later tested, race, cognitive ability, motivation to study, study self-efficacy, number of times a practice test was taken, studying with a tutor or study group, and whether they had set aside certain times to study.

The study concluded that the Black-White standardized mean-score difference on the TPM test was less than half that obtained on a measure of cognitive ability. Also, the amount of time the TPM manual was studied was the primary variable that predicted TPM test performance, with the lowest Black-White mean-score difference on the TPM test being for those who reported studying the TPM manual the most. However, cognitive ability was also a significant predictor of performance on the TPM and cognitive ability was correlated with race.
During this study Black test takers benefited more than White test takers from increased study time during the final two weeks before the TPM test was administered, but this benefit was eliminated when cognitive ability was controlled. Furthermore, those with higher levels of cognitive ability benefited significantly more from increased study times than those with lower levels of cognitive ability.

The TPM test appears to offer a fair and valid method for measuring the cognitively-loaded aptitudes of reading and learning written text. When appropriate it should be considered as an alternative employment practice as specified in Section 3 of the federal Uniform Guidelines on Employee Selection Procedures (1978), if future research finds that it is a valid alternative to cognitive ability testing.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>xi</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>xiii</td>
</tr>
</tbody>
</table>

## CHAPTER

### I. INTRODUCTION

Adverse Impact and Subgroup Differences – Two Sides of the Same Coin? ....3

Adverse Impact and Subgroup Differences in TPM-Style Testing .................4

Hypothesis 1....................................................................................................6

Potential Alternative Explanations for Lower Subgroup Differences During TPM Testing Compared to Cognitive Ability Testing ........................................6

Relationship Between Cognitive Ability and Learning as Measured by the TPM Test............................................................................................................7

Hypothesis 2....................................................................................................9

Relationship of Amount of Time Studying the TPM Manual and Learning as Measured by the TPM Test ................................................................. 9

Hypothesis 3a................................................................................................10

Hypothesis 3b................................................................................................11

A Proposed Model of TPM Test Preparation .................................................11

Relationship of Motivation to Study the TPM Manual and the Amount of Time the TPM Manual was Studied .............................................................13

Hypothesis 4a................................................................................................13
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPM Testing – Psychometric Properties</td>
<td>29</td>
</tr>
<tr>
<td>Reliability of entry-level firefighter TPM testing</td>
<td>29</td>
</tr>
<tr>
<td>Content-related validity of the TPM test used in the current study</td>
<td>29</td>
</tr>
<tr>
<td>Criterion-related validity of trainability testing and of previous</td>
<td>32</td>
</tr>
<tr>
<td>versions of entry-level firefighter TPM testing</td>
<td></td>
</tr>
<tr>
<td>Construct validity of TPM testing in general</td>
<td>34</td>
</tr>
<tr>
<td>Relationship of Trainability Tests and General Mental Ability</td>
<td>35</td>
</tr>
<tr>
<td>Trainability Testing as an Alternative Selection Methodology to Reduce</td>
<td>36</td>
</tr>
<tr>
<td>Impact Associated with Traditional Measures of Cognitive Ability</td>
<td></td>
</tr>
<tr>
<td>Hypothesis 1</td>
<td>39</td>
</tr>
<tr>
<td>Potential Alternative Explanations for Lower Subgroup Differences During</td>
<td>40</td>
</tr>
<tr>
<td>TPM Testing Compared to Cognitive Ability Testing</td>
<td></td>
</tr>
<tr>
<td>Role of Cognitive Ability and TPM Test Performance</td>
<td>41</td>
</tr>
<tr>
<td>Hypothesis 2</td>
<td>42</td>
</tr>
<tr>
<td>Study Time and Performance</td>
<td>42</td>
</tr>
<tr>
<td>Hypothesis 3a</td>
<td>46</td>
</tr>
<tr>
<td>Potential Moderator Effect of Cognitive Ability on Study Time</td>
<td>46</td>
</tr>
<tr>
<td>Hypothesis 3b</td>
<td>48</td>
</tr>
<tr>
<td>Expanded Model</td>
<td>49</td>
</tr>
<tr>
<td>Effect of Motivation on Study Time and Test Performance</td>
<td>49</td>
</tr>
<tr>
<td>Hypothesis 4a</td>
<td>51</td>
</tr>
<tr>
<td>Hypothesis 4b</td>
<td>51</td>
</tr>
<tr>
<td>Hypothesis 4c</td>
<td>51</td>
</tr>
<tr>
<td>Self-efficacy, Learning, Study Time, and TPM Performance</td>
<td>51</td>
</tr>
</tbody>
</table>
Hypothesis 5a........................................................................................................54
Hypothesis 5b........................................................................................................54
Hypothesis 5c........................................................................................................54
Mediating Effect of Cognitive Ability on Race Predicting TPM Test
Performance ..............................................................................................................54
Hypothesis 6..................................................................................................54
Possible Interaction of Race and Studying ............................................................55
Research Question 1 .....................................................................................58
Research Question 2 .....................................................................................59
Study Strategies .....................................................................................................59
Test preparation training/coaching ..................................................................59
Effect of taking practice exams on test performance ....................................64
Research Question 3 .....................................................................................67

III. METHODOLOGY ......................................................................................................68
Participants.............................................................................................................68
Demographic Data .................................................................................................70
Measure of Study Behavior ...................................................................................70
Pre-employment Selection Measures.................................................................74
Test preparation manual test (TPM test)..............................................................75
Cognitive ability test (CAT; predictor variable).....................................................76
Design and Procedure ............................................................................................76

IV. RESULTS ....................................................................................................................79
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Race of Test Takers Who Responded to Characteristics Survey</td>
<td>69</td>
</tr>
<tr>
<td>2.</td>
<td>Demographics of Black and White Participants Who Responded to Characteristics Survey</td>
<td>71</td>
</tr>
<tr>
<td>3.</td>
<td>Means, Standard Deviations, and Intercorrelations Between Measures</td>
<td>80</td>
</tr>
<tr>
<td>4.</td>
<td>Characteristic, Demographic, and Test-Score Differences Between Black and White Participants</td>
<td>81</td>
</tr>
<tr>
<td>5.</td>
<td>Characteristic, Demographic, and Test-Score Differences Between Younger (under 40) and Older (40 and over) Participants</td>
<td>82</td>
</tr>
<tr>
<td>6.</td>
<td>Characteristic, Demographic, and Test-Score Differences Between Male and Female Participants</td>
<td>83</td>
</tr>
<tr>
<td>7.</td>
<td>Mean Effect Size-Score Differences ($d$) Between Black and White Test Takers on the TPM Test and the Measure of $G$.</td>
<td>87</td>
</tr>
<tr>
<td>8.</td>
<td>Regression Analysis Summary for the Interaction of Total Time Studying and Cognitive Ability Predicting TPM Test Score (N = 1336)</td>
<td>90</td>
</tr>
<tr>
<td>9.</td>
<td>Regression Analysis Summary for the Interaction of Final Two Weeks Studying and Cognitive Ability Predicting TPM Test Score (N = 1334)</td>
<td>92</td>
</tr>
<tr>
<td>10.</td>
<td>Regression Analysis Summary for the Interaction of Total Time Studying and Race Predicting TPM Test Score (N = 1336)</td>
<td>102</td>
</tr>
<tr>
<td>11.</td>
<td>Regression Analysis Summary for the Interaction of Final Two Weeks Studying and Race Predicting TPM Test Score (N = 1334)</td>
<td>104</td>
</tr>
<tr>
<td>12.</td>
<td>Mean Effect-Size Differences Between Black and White TPM Test Scores at Various Total Study-Time Intervals</td>
<td>106</td>
</tr>
<tr>
<td>13.</td>
<td>Mean Effect-Size Differences Between Black and White TPM Test Scores at Various Study-Time Intervals During the Final Two Weeks Before the Test Was Administered</td>
<td>107</td>
</tr>
</tbody>
</table>
14. Mean Effect-Size Differences Between Black and White Scores on the Measure of Cognitive Ability at Various Total Study-Time Intervals ..........111
15. Mean Effect-Size Differences Between Black and White Scores on the Measure of Cognitive Ability at Various Study-Time Intervals During the Final Two Weeks Before the Test Was Administered........................................112
16. Adverse Impact on TPM Test Between Black and White Test Takers at Varying Amounts of Total Study Time ...............................................................117
17. Adverse Impact on TPM Test Between Black and White Test Takers at Varying Amounts of Study Time During the Final Two Weeks Before the Tests Were Administered.....................................................................................119
18. Regression Analysis Summary for Study Strategy Characteristics Predicting TPM Test Scores When Holding Total Time Studying and Cognitive Ability Constant (N = 1290) ...........................................................................................................................................122
19. Regression Analysis Summary for Study Strategy Characteristics Predicting TPM Test Scores When Holding Time Studying During the Final Two Weeks and Cognitive Ability Constant (N = 1302)..................................................................123
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Simple proposed two-factor model for predicting TPM test performance</td>
<td>8</td>
</tr>
<tr>
<td>2.</td>
<td>Conceptual model of components related to performance on an entry-level firefighter TPM test</td>
<td>12</td>
</tr>
<tr>
<td>3.</td>
<td>Possible contribution of study time by race interaction to TPM test scores</td>
<td>18</td>
</tr>
<tr>
<td>4.</td>
<td>TPM scores of low, mean, and high cognitive ability participants as a function of the total amount of time studying the TPM manual</td>
<td>91</td>
</tr>
<tr>
<td>5.</td>
<td>TPM scores of low, mean, and high cognitive ability participants as a function of the final two weeks studying the TPM manual</td>
<td>94</td>
</tr>
<tr>
<td>6.</td>
<td>Black and White participants' TPM test scores as a function of the amount of time studying the TPM manual during the final two weeks before the test was administered</td>
<td>105</td>
</tr>
<tr>
<td>7.</td>
<td>Mean effect-size differences ($d$) in TPM test scores as a function of the total number of hours studying the TPM manual before the test was administered</td>
<td>109</td>
</tr>
<tr>
<td>8.</td>
<td>Mean effect-size differences ($d$) in TPM test scores as a function of the hours studying the TPM manual the final two weeks before the test was administered</td>
<td>110</td>
</tr>
<tr>
<td>9.</td>
<td>Mean effect-size differences ($d$) in scores on the measure of cognitive ability as a function of the total number of hours studying the TPM manual before the test was administered</td>
<td>113</td>
</tr>
<tr>
<td>10.</td>
<td>Mean effect-size differences ($d$) in scores on the measure of cognitive ability as a function of the hours studying the TPM manual the final two weeks before the test was administered</td>
<td>114</td>
</tr>
<tr>
<td>11.</td>
<td>Simple two-factor model for predicting TPM test performance</td>
<td>127</td>
</tr>
<tr>
<td>12.</td>
<td>Selected aspects of a model of the relationships (beta) between components related to performance on an entry-level firefighter TPM test</td>
<td>132</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

Measures of cognitive ability have consistently been shown to be among the best predictors of job performance for a wide variety of occupations (e.g., Hunter & Hunter, 1984; Schmidt & Hunter, 1981, 1996, 2000), primarily because intelligent people learn job knowledge faster and learn more of it (Schmidt & Hunter, 2000). However, it has been suggested by many researchers, including Murphy (2002) and Outtz (2002), that the relatively high levels of adverse impact associated with cognitive ability measures make it difficult to defend their use in high stakes employment decisions such as selection for safety forces jobs in the public sector (Murphy, Cronin, & Tam, 2003).

One possible solution to the problem of high levels of adverse impact associated with cognitive ability testing is to simply eliminate such tests from selection batteries. However, public-safety organizations that have attempted to implement this option have encountered negative results. For example, following the elimination of the use of cognitive ability testing in 1987 by the Washington, DC police department, the failure rate at the Washington, DC police academy rose to more than 80% (Carlson, 1993; Schmidt & Hunter, 2000). As a result, the rigor of the academy’s training was compromised and the solution rate for homicides moved from one of the highest in the nation to one of the lowest (Schmidt & Hunter, 2000).
A second alternative is a compensatory approach (e.g., Hough, Eaton, Dunnette, Kamp, & McCloy, 1990) where cognitive ability measures are combined with scores from non-cognitive measures, such as personality tests, which typically have low levels of adverse impact. However, Ryan, Ployhart, and Friedel (1998) found that this approach increased adverse impact against women, even as it decreased adverse impact against minorities. Furthermore, other studies suggest that this alternative may have a limited impact on the reduction of adverse impact (e.g., Barrett, Kramen, & Lueke, 2003; Bobko, Roth, & Potosky, 1999; Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997).

A third alternative is the use of work sample or trainability tests. Such tests allow the interactions of the test’s various components to occur “naturally” (Asher & Sciarrino, 1974; Smith, 1991). This natural interaction would be expected to result in a better prediction of future performance than a series of tests that individually measured job-related components, while still reducing adverse impact.

The current study examined the use of a form of trainability testing, called TPM (for Test Preparation Manual) testing, which was designed to simulate the learning of job-related information for the job of a firefighter. During trainability testing, a person is assessed as to how well he or she can learn the knowledge and behaviors required to perform a job (Robertson and Downs, 1979). In TPM testing, applicants for a firefighting position study a manual containing information pertinent to the job of firefighting (TPM manual) and then, during the testing session, answer questions from memory about information contained in the manual (Campbell, 1982).

The TPM testing process has been proposed as a method of reducing adverse impact by allowing test takers with lower levels of ability to compensate by increasing
the amount of time they study the material (Biddle & Associates, 1997). In addition, TPM testing measures the outcome of a combination of the amount of time spent studying and the efficacy of such studying, which is likely to reflect both motivational factors and cognitive ability. If we assume that diverse groups may differ in cognitive ability but not in motivation, the combination of motivation and cognitive ability should “naturally” reduce adverse impact (Smith, 1991).

In the current study, it was assumed that the TPM test would result in less adverse impact than a test of cognitive ability. However, given the lack of literature on this topic, this assumption was treated as the first hypothesis to be tested. Second, if the TPM does result in less adverse impact, this would leave open the question as to why this occurred. Was it because TPM testing allowed for a “natural interaction” of cognitive ability and motivation? Thus, a second purpose was to determine if the factors impacting performance on the TPM test could be identified and if this would increase our understanding of the trainability testing paradigm.

Adverse Impact and Subgroup Differences – Two Sides of the Same Coin?

Technically the issue of adverse impact and subgroup mean effect-size differences are related but separate issues. However, most of the extant literature fails to distinguish between the two and has often used the term “adverse impact” to refer to the size of the standardized subgroup mean difference.

The Federal Uniform Guidelines on Employee Selection Procedures (1978) used the term “adverse impact” to describe substantially different rates of selection when employment decisions are made. These selection differences are typically the result of two factors. First, selection rate differences frequently depend on where the pass/fail
cutoff point is set. As cutoff scores are set higher, the level of adverse impact typically increases (Wollack, 1994). Second, the size of the differences in the test scores between subgroups can also play a role in differences in selection rates. These test-score differences are typically expressed as mean subgroup effect-size differences, which are often referred to as $d$.

While adverse impact and subgroup differences are sometimes used interchangeably for describing potential differences in the literature, they can lead to dramatically different conclusions. For example, tests without any mean subgroup effect-size difference can still have adverse impact if the variance in the test scores of each of the subgroups is sufficiently different. For this reason, both the Black-White subgroup mean effect-size difference and potential adverse impact against Black test takers were examined in the current study. Finally, it should also be noted that the term adverse impact is used throughout this document to describe substantially different selection rates, whereas the term subgroup differences is used to describe mean effect-size differences.

*Adverse Impact and Subgroup Differences in TPM-Style Testing*

Adverse impact against minorities has been demonstrated to be relatively low for trainability tests in general (Higuera & Riera, 2004). Furthermore, it has been reported that TPM type testing results in less adverse impact than some other types of written tests. For example, Campbell (1982) examined several fire departments that were in relatively close geographical proximity to one another in Northern California. He reported that there was no violation of the 80% adverse-impact rule for any of the five fire departments that used TPM testing for selecting their firefighters. Conversely, there
were 80% rule violations for all of the other ten fire departments he examined that used what he referred to as entry-level firefighter “aptitude tests” for selecting entry-level employees.

Unfortunately, there appears to be only one published study regarding TPM testing that resulted in the ability to calculate Black-White mean effect-size differences. In 1996, Benver reported on a TPM test that was used for selecting entry-level police officers in Kentucky. During this selection process test takers were allowed to read and study the text for months in advance of the test being administered. An examination of Benver’s data found a Black-White standardized mean difference of $d = .734$ (Black $n = 126$, White $n = 536$), which appears to be relatively high for a practice that is being considered as an alternative to cognitive ability testing although it is still lower than the one standard deviation difference commonly reported for cognitive ability. Even so, the mean difference reported by Benver is still lower than the Black-White GAES of $d = 1.0$.

A number of unpublished subgroup-difference analyses have been conducted on the firefighter TPM used in the current study. Mean subgroup differences of $d = .30$ to $.60$ have been found during the administration of the first seven versions of the TPM tests, which were extremely similar to the eighth version used in the current study (S. Bell, personal communication, February 22, 2004).

As previously indicated, it was an assumption of this study that the TPM would have less adverse impact than a cognitive ability test. Still given the lack of literature and research on this question, it will be confirmed as a part of testing the first hypothesis:
Hypothesis 1. The Black-White subgroup mean difference on the measure of learning, the TPM test, will be smaller or less than the subgroup mean difference on cognitive ability.

Potential Alternative Explanations for Lower Subgroup Differences During TPM Testing Compared to Cognitive Ability Testing

The question might be asked – why would the TPM test result in lower subgroup differences and less adverse impact than a measure of cognitive ability? There appear to be two possibilities. The first might be that the TPM test scores contained more random error than cognitive ability scores. However, previous TPM test administrations for entry-level firefighter selection have resulted in reliability coefficients in the range of .90, which is similar to that for cognitive ability testing.

Furthermore, if the TPM resulted in lower levels of adverse impact primarily because it was capturing high levels of random error, then it would be unlikely that the TPM would be a valid predictor of firefighter job performance. In addition, in order to consider TPM testing as an alternative to cognitive ability testing, we need to have some evidence that it is a valid predictor of job performance, or in this case firefighter performance.

Previous research has demonstrated that job performance can be effectively predicted by job knowledge (e.g., Hunter & Schmidt, 1996; Palumbo, Miller, Shalin, & Steele-Johnson, 2005; Schmidt & Hunter, 1998). Interestingly, a study by Palumbo, Miller, Shalin, and Steele-Johnson found that a job knowledge measure similar in construction to a trainability test accounted for significantly greater variance in predicting simulated job performance than a cognitive ability measure (26% v. 12%). This is an
extremely important finding since the TPM test was designed to measure a person’s ability to learn declarative job knowledge.

There is empirical data supporting the relationship between TPM testing and fire academy performance including a 1982 study, which found an uncorrected correlation of 0.54 between TPM test scores and fire academy mid-term grades for members of two fire academy classes in a southern California city ($p < .01$, two-tail, $n = 14$ and $n = 16$; Biddle & Associates, 1997). Thus, we have evidence, albeit limited, that the TPM is both a valid test and a reliable test. It seems unlikely that the reduction in adverse impact would simply be the result of increased random error.

The second possibility was that factors other than cognitive ability, such as the amount of time a person studied the TPM manual and motivational beliefs, might have contributed to the prediction of TPM scores in such a way that this natural interaction resulted in reduced subgroup differences on the TPM test. As a first step, a simple model of the effect of cognitive ability and study time on the learning measure, the TPM test, was proposed and appears in Figure 1.

**Relationship between Cognitive Ability and Learning as Measured by the TPM Test**

According to the proposed simple model, cognitive ability should predict TPM test performance, which is basically a learning measure. Studies (e.g., Hunter & Hunter, 1984; Schmidt, Hunter & Outerbridge, 1986; Schmidt & Hunter, 1992) have consistently shown a positive relationship between cognitive ability and the learning of job knowledge. More specifically, Schmidt and Hunter (1998) reported cognitive ability had an average predictive validity of .56 “for performance (amount learned) in job training
Figure 1. Simple proposed two-factor model for predicting TPM test performance.
programs” (p. 266). Since the TPM test was designed as a method to measure job-related learning, the following prediction was made:

_Hypothesis 2._ Cognitive ability will be positively related to learning as measured by the TPM test.

_Relationship of Amount of Time Studying the TPM Manual and Learning as Measured by the TPM Test_

It seemed logical that those who studied the TPM manual longer would perform better on a test containing questions about that manual. However, research concerning the relationship between study time and performance, especially school performance as determined by grades, has been equivocal. For example, some studies (e.g., Blustein, Judd, Krom, Viniar, Padilla, Wedemeyer, & Williams, 1986; Cheung & Kwok, 1998; Michaels & Miethe, 1989) have reported finding positive relationships between study time and grades, others (e.g., Schuman, Walsh, Olson, & Etheridge, 1985) have reported finding no significant relationship, and still others (e.g., Greenwald & Gillmore, 1997; Olivares, 2002) have even reported finding negative relationships.

A search of the literature and other sources revealed two instances that had specifically examined the relationship between TPM test performance and studying of the manual (i.e., D. A. Biddle, personal communication, October 20, 2003; Benver, 1996). These reports, which are examined in more detail in Chapter 2, found that job applicants who indicated they had used the TPM manual more frequently performed better on a TPM test than those who reported using the TPM manual less frequently. A similar finding was expected during the current study, leading to a hypothesis regarding study time and TPM performance:
Hypothesis 3a. The self-reported amount of time spent studying the TPM manual will be positively related to TPM test scores.

Next, it appeared possible that the relationship between self-reported time studying the TPM manual and TPM performance might be moderated by cognitive ability. Baron and Kenny (1986) suggested that moderator variables are most likely to operate when there is a weak or inconsistent association between a predictor (e.g., amount of time studying) and a criterion (e.g., grades). As shown above, the relationship between the amount of time studying and performance has been found to be inconsistent.

In addition, Olivares (2002) found that cognitive ability, which was positively related to school grades in his study, moderated the relationship between the amount of time studying and school grades. Furthermore, he also found that the amount of time studying was inversely related to cognitive ability, which he explained as those with the lowest levels of cognitive ability attempting to compensate through additional time studying.

Why is this important? Some learning models such as the Interactive, Compensatory Model of Learning (ICML; Schraw, Brooks, & Crippen, 2005) and the Compensatory-Encoding Model (C-EM; Walczyk, 1993, 2000) postulate that a test taker’s weaknesses in some areas could be compensated for through strengths or strategies in other areas. Of course, this compensation would only be effective for reducing mean effect-size differences if the amount of time studying was more beneficial to those with lower levels of ability than those with higher levels of ability.

Walczyk (personal communication, November 7, 2005) has offered a suggestion as to why this might be so. He indicated that when people with higher levels of ability
study for long periods of time they have a tendency to become bored. Whereas those with lower ability levels can become more engaged as they study longer. Thus, studying for a longer period of time might not be equally beneficial to those who possess high and low ability levels when there is a finite amount of information to be learned. This would allow those with lower ability levels, in effect, to “catch up” to those with higher ability levels. This leads to the following hypothesis:

Hypothesis 3b. The effect of self-reported study time on TPM test scores will be moderated by cognitive ability, such that study time will result in higher TPM scores for those lower in cognitive ability as compared to those higher in cognitive ability.

A Proposed Model of TPM Test Preparation

It was proposed in Hypothesis 2 that cognitive ability would have a positive, significant relationship to learning as measured by the TPM test. It was further proposed in Hypotheses 3a that there would be a significant, positive relationship between the amount of time the TPM manual was studied and scores on the TPM test. Finally, it was proposed in Hypothesis 3b that cognitive ability would moderate the relationship between the amount of time studying and TPM test performance.

In addition, it appeared likely that there would be other characteristics that were likely to influence the amount of time the job candidates studied the TPM manual and that they would also influence performance on the TPM test. A graphic model of the factors proposed to be related to time studying the TPM manual and TPM test performance is presented in Figure 2. Figure 2 is based on the expectation that race, motivational beliefs, cognitive ability, and the amount of time studying the TPM manual
Figure 2. Conceptual model of components related to performance on an entry-level firefighter TPM test.

Note: The *Time studying the TPM manual* variable reflects two different time periods: The total time spent studying the TPM manual and the time spent studying the TPM manual the final two weeks before the test was administered, which were both expected to act similarly. Solid arrows indicate a direct relationship between variables. The dashed arrow indicates a hypothesized moderator effect of cognitive ability on the relationship between time studying the TPM manual and TPM test performance.
will all be related to TPM test performance. An examination of several particular aspects of the theoretical model presented in Figure 2 will be made during the current study.

Relationship of Motivation to Study the TPM Manual and the Amount of Time the TPM Manual was Studied

The relationship between motivation and performance has been studied extensively throughout the literature. Furthermore, most of the influential models of performance include motivation as an essential factor (e.g., Campbell, McCloy, Oppler, & Sager, 1993; Motowidlow, Borman, & Schmit, 1997). More specifically, the positive relationship between motivation and learning has been extensively explored and tested (e.g., Dweck, 1986; Eppler & Harju, 1997).

Icek and Ajzen (1988) postulated that behaviors are impacted by motivational factors. This results in what they call the “intention to engage in a certain behavior,” such as how willing people to attempt a behavior and how much effort they will be exerting (p. 113). They further indicated that research evidence supported the concept that intentions are closely related to overt volitional behavior. Thus, it appeared likely that those who indicated they were more motivated to study the TPM manual in the current study would in fact spend more time studying it, leading to the following prediction:

Hypothesis 4a. Self-reported levels of motivation to study the TPM manual will be positively related to the self-reported amount of time spent studying.

Relationship of Motivation to Study the TPM Manual and Learning as Measured by the TPM Test

As indicated by Hypothesis 4, it was anticipated that there would be a positive relationship between motivation to study the TPM manual and the amount of time it was
studied. Also, Hypothesis 3a predicted that there would be a positive relationship between the amount of time studying the TPM manual and TPM test performance. A logical extension of these two hypotheses was proposed.

*Hypothesis 4b.* Self-reported levels of motivation to study will be positively related to TPM test scores.

In the current instance, it appeared logical that motivation to study the TPM manual was not likely a direct cause of better performance on the TPM test. Instead, it was thought that it was primarily through studying the TPM manual that this effect occurred. Since it was anticipated that motivation to study the TPM manual would have a direct effect on the amount of time the TPM manual was studied, and that the amount of time studying the TPM manual would have a direct effect on TPM test performance, the following hypothesis was proposed:

*Hypothesis 4c.* The effect of motivation to study on TPM test scores will be mediated by study time.

**Relationship of Study Self-Efficacy and the Amount of Time the TPM Manual was Studied**

Self-efficacy is a core self-evaluation trait that has been found to be related to performance (Judge & Bono, 2001). Self-efficacy refers to the subjective probability that a person is capable of successfully executing a task (Bandura, 1977, 1982). To this end, some studies have reported that people with higher levels of perceived self-efficacy are more likely to persist longer, use more regulatory and deep-level strategies, and achieve higher grades than those who are unsure of their ability to perform (e.g., Bandura, 1997; Gist & Mitchell, 1992; Pajares, 1996).
Furthermore, Schunk (2003) indicated that an individual’s self-efficacy promotes learning, such that higher levels of self-efficacy would be related to better performance. More specifically in regards to the current study, Ryan, Ployhart, and Friedel (1998) reported self-efficacy was positively related to ability-test performance for an entry-level firefighter position.

In the current study, self-efficacy was measured by a single item designed to elicit information concerning the extent to which the test takers agreed that the more time they spent studying the TPM manual would lead to their performing better on the TPM test. Based on the previous research which has consistently reported a positive relationship between self efficacy and performance, the following hypothesis about study self-efficacy was offered.

*Hypothesis 5a.* Self-reported levels of study self-efficacy will be positively related to the self-reported amount of time spent studying.

*Study Self Efficacy and TPM Test Performance*

As with motivation to study, it was also anticipated that there would be a direct positive relationship between study self-efficacy and the amount of time the TPM manual was studied. Furthermore, it was anticipated that there would be a positive relationship between test takers’ self-reported levels of study self-efficacy and TPM test performance. With that in mind the following hypothesis concerning study-self efficacy and learning as measured by the TPM test was offered.

*Hypothesis 5b.* Self-reported study self-efficacy will be positively related to TPM tests.
As with motivation to study, it appeared logical that higher levels of study self-efficacy in and of itself were likely not a direct cause of better performance on the TPM test. Instead, it appeared that a relationship between study self-efficacy and TPM test performance would take place primarily through the studying of the TPM manual. Thus, the following hypothesis was offered:

*Hypothesis 5c.* The effect of study self-efficacy on TPM test scores will be mediated by study time.

*Relationship of Race and Learning as Measured by the TPM Test*

Researchers have indicated that mean-score differences between Black and White test takers are more likely to occur when using more heavily cognitively-loaded devices than when using those that are less cognitively-loaded (e.g., Spearman, 1904, 1927; Jensen, 1985). Nyborg and Jensen (2000) went so far as to say that Black-White differences on cognitive ability “should no longer be regarded as just a hypothesis but as an empirically established fact” (p. 599).

Since the TPM is likely to be related to cognitive ability, it could be inferred that a relationship between race and performance on the TPM test would likely be mediated by differences in cognitive ability. Hypothesis 6 was offered to test the potential mediating effect of cognitive ability on race when predicting TPM test performance.

*Hypothesis 6.* The effect of race on TPM test scores will be mediated by cognitive ability.

*Relationship between Race, Study Amount, and TPM-test performance*

One of the main focuses of the current study was to determine whether Black job applicants benefited more from increased time studying the TPM manual than White job
applicants. In other words, it attempted to answer the question of whether the slope of the relationship between study amount and performance would be significantly different for Black as compared to White job applicants.

As mentioned earlier, both the ICML theory (Schraw, et al., 2002) and the C-EM theory (Walczyk, 1993, 2000) postulated that increased time spent reading and/or studying can allow test takers to use multiple strategies, including increased preparation time, to compensate for lower levels of ability. Furthermore, Walczyk (personal communication, November 7, 2005) postulated that those with lower ability levels would likely benefit more from increased time studying than those with higher ability levels.

Following the logic put forth by Walczyk, since Black test takers typically perform lower on measures of cognitive ability than White test takers, it was anticipated that Black test takers would benefit more from an increased amount of time studying a finite amount of study material than White test takers (see Figure 3). However, as mentioned previously, relationships between the amount of time studying and performance have been inconclusive during many studies. To help examine these relationships, the following research question was asked:

*Research Question 1.* Will time spent studying and race interact to effect learning, as measured by the TPM test, such that Black test takers benefit more from increased study time than White test takers?

Next, if Black test takers benefited more from an increased amount of time studying the TPM manual then it appeared likely that the lowest mean effect-size differences and the least amount of adverse impact when using the TPM test would result
Figure 3. Possible contribution of study time by race interaction to TPM test scores.
when the reported study amount was highest for both Black and White test takers. This led to the following research question being asked:

*Research Question 2.* Will the Black-White subgroup mean difference and potential adverse impact on the TPM test decrease as the amount of study time increases?

**Effect of Other Individual-Difference Characteristics on the Relationship Between the Time Studying the TPM Manual and TPM Test Scores**

Information about test-taker behaviors and characteristics, other than study amount, motivation, study self-efficacy, and race were also examined during the current study. For example, just prior to taking the TPM and cognitive ability tests, participants were asked to provide responses to survey items regarding how many times they may have taken a practice test which was offered, whether they prepared with a study group or a tutor, and whether they set aside certain times to study the TPM manual. To investigate this, the following research question was tested:

*Research Question 3.* A combination of study techniques (including setting aside certain times to study, studying with a tutor or study group, and the number of times a practice test was taken) will explain variance beyond that explained by study time.

**Contributions of the Current Study**

This study offers a significant contribution to the literature in several ways. It expands on Benver’s (1996) finding that entry-level law enforcement job applicants who more frequently used a TPM manual performed better on its related written test ($r = .304$, $p < .01$, two-tailed, $n = 555$). However, no data was collected by Benver pertaining to the test administrations he studied. Thus, it remained unknown whether those who studied more and scored higher may have also had higher levels of cognitive ability, which might
have served as an alternative explanation for their higher test scores. To address this issue, the current study examined whether the time spent studying as well as other factors are positively related to TPM test scores when cognitive ability is held constant.

While preparation for academic tests (e.g., Smyth, 1989; Snedecor, 1989; Whitla, 1988) and pre-employment tests (e.g., Ryan, Ployhart, Greguras, & Schmit, 1998) have been studied in the context of attendance at preparation training sessions, a review of the literature examined in Chapter 2 revealed a dearth of studies that have directly examined the relationship between the amount of actual test-preparation time (i.e., reading and studying) and test performance in a pre-employment setting. To better understand this issue, the current study directly tests whether test-taker characteristics, such as the amount of time spent studying and motivation, were positively related to trainability test performance when controlling for cognitive ability. The finding of a significant positive relationship when cognitive ability was held constant would help to support the theory that trainability tests can add incremental validity by measuring job-related attributes beyond cognitive ability.

The current study also presents an alternative to mechanically combining scores from separate measures of cognitive and non-cognitive abilities when attempting to reduce adverse impact against some protected groups of test takers. Cognitive and non-cognitive measures are frequently combined mechanically, without any theoretical underpinning and often without empirical justification. The aforementioned alternative to compensatory scoring consists of a single job-related and theory-driven measure (i.e., TPM test) on which performance is the result of cognitive abilities and aptitudes (such as reading and remembering written information) that are affected by non-cognitive
characteristics (such as the amount of time spent studying, confidence, and motivation) in a compensatory fashion.

The final contribution relates to Ryan et al.’s (1998) and Sackett, Burris, and Ryan’s (1989) concern that non-equivalent control groups in their studies prevented them from examining the study preparation efforts of those who failed to attend preparation sessions. Virtually all test takers who agreed to participate in the research that provided the archival data for the current study provided information about their test-preparation behaviors and test-related characteristics, such as amount of time spent studying the TPM manual, taking the practice test and their level of confidence, motivation, and prior academic achievement. By using this approach, results from both those who reported preparing diligently for the TPM test and those who reported not preparing at all were directly compared.

Conclusion of Chapter One and Overview of Future Chapters

The present chapter has presented the conceptual model of the relationship between study time and test performance on a TPM-style test and has offered predictions related to this relationship. In the second chapter, the relevant literature concerning how study time and other test-taker behaviors/characteristics might influence TPM test performance, especially as it is related to race, is reviewed. The methodology related to the assessing the hypotheses will be described in chapter three. In the fourth chapter, descriptive statistics and findings from the tests of the hypotheses will be reported. Finally, the fifth chapter will include a discussion of the findings as well as the limitations of the study and the implications for future research in this area.
CHAPTER II
LITERATURE REVIEW

The primary purpose of the current study was to examine a form of trainability test that had been purported to reduce adverse impact when compared to cognitive ability tests. Cognitive ability testing has been shown to be one of the best predictors of job performance (e.g., Hunter & Hunter, 1984; Schmidt & Hunter, 1981, 1998, 2000). However, cognitive ability testing typically results in mean score differences against some subgroups of job applicants, which contributes to adverse impact (e.g., Murphy, 2002; Outtz, 2002). Selection practices that result in adverse impact can sometimes result in litigation by those who are negatively affected (Biddle, Kuthy, & Nooren 2003).

Although adverse impact litigation can occur with any type of occupation or job, it has been especially prevalent in the public sector when selecting for protective service jobs (Biddle, Kuthy, & Nooren 2003). Terpstra and Kethley (2002) indicated that protective service jobs, such as firefighters and peace officers, are especially likely to be targets for litigation. Specifically, they discovered that litigation against protective service employers accounted for 75 percent of service-job court cases, even though these positions accounted for only 12 percent of all workers in service jobs. One of the predominate factors cited in public safety-related litigation is adverse impact in the selection process (Guion, 1998). Therefore, efforts aimed at minimizing the amount of adverse impact associated with selection devices may help to reduce litigation associated
with entry-level selection (Biddle, Kuthy, & Nooren, 2003). It is for this reason that the current study was conducted in the context of a protective service job of entry-level firefighter.

In addition, Biddle et al. (2003) indicated that candidates who perceive that the selection process is “face valid” may be less likely to pursue litigation. Face validity typically refers to devices that have a transparent relationship to the job for which the applicant or candidate is being tested. “Face valid” selection devices which also minimize adverse impact are therefore highly desirable in litigation-prone public sector employment settings. Trainability testing, which measures a person's ability to learn and perform job-related tasks, is a “face valid” method that holds the promise of decreasing adverse impact during pre-employment testing.

This chapter reviews pertinent aspects of the literature related to the predictions and questions that were offered during the current study. It begins with a review of some of the highlights in the literature related to cognitive ability testing. A review of cognitive ability testing is presented first since this type of testing has been consistently related to the largest subgroup differences in mean scores and adverse impact when compared to other types of selection testing. That is followed by a review of the literature relating to trainability testing; specifically focusing on TPM testing. TPM testing has been proposed by some as an alternative employment practice that purportedly results in lower adverse impact than cognitive ability testing. More specifically, this section of the chapter focuses on the potential influence of the amount of time studying on TPM test performance. That is followed by a review of two motivational belief variables, motivation and self-efficacy, which may offer additional explanation as to why there might be lower subgroup
differences during TPM testing than during measures of cognitive ability. Next, a review of the literature regarding the relationship between race, cognitive ability, and TPM test performance is offered. This is followed by a review of the literature related to the influence of race on the relationship between study time and TPM test performance, especially as this relates to potentially lower subgroup differences. Finally, a review is offered of how study strategies, such as taking a practice test, studying with a tutor or study group, or setting aside certain time to study the TPM manual are likely to be related to TPM performance.

*Cognitive Ability Testing*

Cognitive ability has been used as a predictor of job success for more than 80 years (Viswesvaran & Ones, 2002). More specifically in regards to the current study, Terman (1917) reported in the first edition of the *Journal of Applied Psychology* that general intelligence was one of the most important qualities needed to be a successful firefighter. To this end, Terman used the *Stanford-Binet Intelligence Scale* for testing both police and firefighter applicants in San Jose, California in 1916.

Discovered by Spearman in 1904, *g* is the highest-order common factor that can be extracted from a battery of diverse tests of assorted cognitive abilities using hierarchical factor analysis (Jensen, 1999). Evidence for a single factor was demonstrated by numerous early studies, beginning in the first half of the twentieth century (e.g., Holzinger, 1936; Thurstone & Thurstone, 1941). However, even though a single factor had been demonstrated, Carroll (1993) reported that tests currently being used can identify over 70 different cognitive abilities. Neisser et al. (1996) indicated that the overall pattern of correlations between various cognitive measures can generally be
described as being produced by individual differences in \( g \) plus differences in the specific abilities sampled by particular tests.

Neisser et al. (1996) pointed out that there had not been total agreement on the meaning of \( g \) in the intelligence literature. For example, some described it as a measure of neural processing speed (Reed & Jensen, 1992), a statistical finding or mathematical artifact (Gould, 1996; Thomson, 1939), a version of mental energy (Spearman, 1927), or generalized abstract reasoning ability (Gustafsson, 1984). Gottfredson (2002) indicated that “there is much yet to be learned about the nature of \( g \)” (p. 29). Neisser et al. (1996) indicated that the most influential approach to the conceptualization of \( g \) thus far was based on psychometric testing. Their comprehensive overviews of the intelligence literature pointed out the many differences and similarities in the various approaches psychometricians had used to study the measurement of \( g \), along with descriptions of some of the theoretical basis for the various systems of measurement.

It should be noted that not everyone agrees with the idea that there is a single \( g \) factor to intelligence. For example, Cattell (1971) attempted to discredit Spearman’s (1904; 1927) single \( g \) factor primarily because he concluded it was not supportable within a rotated factors theory. His rejection of a single \( g \) factor continues to be postulated primarily by Horn (1998). Cattell (1941; 1963) posited two second-order intelligence factors which he called fluid \((G_f)\) and crystallized \((G_c)\) general intelligence.

Schmidt and Hunter (1998) summarized 85 years of personnel selection research when they said “the most well-known conclusion from this research is that, when hiring employees without previous job experience, the most valid predictor of future performance and learning is general mental ability” (p. 262). A plethora of meta-analyses
examining the use of cognitive ability testing for personnel selection appear to support this claim (e.g., Hartigan & Wigdor, 1989; Hirsch, Northrop, & Schmidt, 1986; Hunter & Hunter, 1984; Levine, Spector, Menon, Narayanan, & Cannon-Bowers, 1996; Schmidt, Gast-Rosenberg, & Hunter, 1980; Vinchur, Schippman, Switzer, & Roth, 1998).

**Adverse Impact When Using Cognitive Ability Measures**

Even though measures of pure cognitive ability typically have among the highest levels of criterion-related validity when predicting job performance in most occupations (Hunter & Hunter, 1984; Schmidt & Hunter, 1981), they also have the highest levels of adverse impact against some minority groups (Campbell, 1996; Gottfredson, 1988). Previous research shows a strong relationship between race and intelligence (Lynn, 2003), with racial group differences during ability testing being examined for over a century (e.g., Galton, 1892, Thorndike, 1921). These examinations of racial group differences in cognitive ability have historically focused on those between Black and White participants (Roth, BeVier, Bobko, Switzer, & Tyler, 2001). Following in this tradition, the current study primarily focused on Black-White differences. It should be noted that the generally accepted mean-score difference between Blacks and Whites on measures of cognitive ability is often about one standard deviation (Hunter & Hunter, 1984; Roth, BeVier, Bobko, Switzer, & Tyler, 2001).

Spearman’s Hypothesis (Spearman, 1904, 1927), later dubbed the “Spearman-Jensen hypothesis” (Osborne, 1980), stated that Black-White average differences were more likely to occur in more heavily g-loaded tests involving complex mental operations, and studies have strongly supported this hypothesis (Jensen, 1985). Nyborg and Jensen (2000) went so far as to say that Spearman’s original theory about Black-White
differences on \( g \) “should no longer be regarded as just a hypothesis but as an empirically established fact” (p. 599).

Previously Considered Methodologies for Reducing Adverse Impact When Using Measures of Cognitive Abilities

Schmidt (2002) asserted that there were two ways that group differences would be reduced when using measures of general cognitive ability: (1) by reducing the reliability of the measure or (2) by assessing more than \( g \) alone. One proposed method to reduce adverse impact by assessing more than \( g \) also entailed measuring personality characteristics and then increasing the weights of personality measures in a selection system in order to mitigate adverse impact from the cognitive ability tests (Ryan, Ployhart, & Friedel, 1998; Sackett & Ellingson, 1997). Schmidt (2002) pointed out that the combination of these predictors led to higher validity and practical utility when compared with single predictors.

However, combining predictors often sacrifices the ability to accurately predict the performance of some important job tasks (Hattrup & Rock, 2002). Furthermore, the Society for Industrial and Organization Psychology Principles for the Validation and Use of Personnel Selection Procedures (2003) warned that attempts to reduce sub-group differences for one group may increase differences in another group. Ryan, Ployhart, and Friedel (1998) argued that the addition of personality measures that do not have adverse impact against minority groups when selecting public-safety officers may introduce adverse impact against women. This is especially troublesome because women are typically under-represented in the firefighting profession (Rosell, Miller, & Barber, 1995).
More critically, the combination of cognitive ability tests and non-cognitive tests may not reduce adverse impact or may have only a limited effect on adverse impact. For example, Sackett and Ellingson (1997) indicated that composites of cognitive and non-cognitive measures can lead to increased or decreased group differences, depending on the circumstances. Similarly, Barrett, Kramen, and Lueke (2003) pointed out that combining scores on a personality test with a measure of cognitive ability will not necessarily result in a decrease of adverse impact. Schmitt, Rogers, Chan, Sheppard, and Jennings (1997) found that when cognitive ability measures were included in a composite along with non-cognitive measures, a violation of the four-fifths rule was still likely to occur. A meta-analysis by Bobko, Roth, and Potosky (1999) found that even non-cognitive composites could violate the four-fifths rule when the selection ratio was less than or equal to 50%. In addition to these cautions, some test developers weigh cognitive and non-cognitive (i.e., personality) measures equally so that someone who scores high on a personality scale and low on a cognitive ability scale would perform just as well on a particular job as someone who scores low on a personality scale and high on a cognitive ability scale. Such a compensatory approach may be questioned given that certain levels of cognitive ability or personality may be required for specific jobs or tasks.

Alternatively, some researchers (e.g., Asher & Sciarrino, 1974; Smith, 1991) have postulated that a work-sample style testing process would allow the interactions of a test’s various components to occur “naturally,” and thus be more likely to elicit behaviors that reflect a person’s true work habits than combinations of tests of specific aptitudes. The TPM test appeared to be an appropriate and valid selection device that allowed for
the natural interaction of characteristics such as motivation, self-efficacy, effort, and ability.

TPM Testing – Psychometric Properties

Trainability tests are samples of work that include a period of learning prior to the performance of the tasks being tested (Higuera & Riera, 2004). The TPM test examined in the current study fits this description since job applicants were instructed to study and learn firefighting-related information. This information is similar to how related information would be learned during training and on the job, and thus candidates would be allowed to demonstrate their ability to memorize, recall, and apply information. This section briefly reviews the literature in regards to the psychometric properties of trainability testing, including the psychometric properties of the type of TPM test used in this study.

Reliability of entry-level firefighter TPM testing. Since reliability limits validity and decreases adverse impact, it was important that the reliability of the TPM test used during the current study was acceptable for testing purposes. Biddle (1996) reported that the reliability of a very similar entry-level firefighter TPM test used in 1985 was .92, which the U.S. Department of Labor (Testing and Assessment, 2000) indicated was "excellent" (page 3-3). Reliability levels of .90 and higher had been generally reported during the administration of the previous versions of the TPM test prior to the version used in the current study being developed (Stacy Bell, personal communication, April 12, 2006).

Content-related validity of the TPM test used in the current study. The Principles for the Validation and Use of Personnel Selection Procedures (Society for Industrial and
Organizational Psychology, 2003) indicated that a content-related strategy for validating selection devices must include analyses that compare the match between test content and (1) work behavior, (2) worker requirements, or (3) outcomes of the work behaviors (page 6). Asher (1972) and Asher and Sciarrino (1974) pointed out that the point-to-point correspondence between the criterion and predictor space contributed to the predictive power of testing. In other words, the theoretical basis for the high validity of trainability tests is that the test itself represents a sample or subset of the criterion domain. This is likely why trainability tests have been shown to possess relatively high levels of validity.

A search of the O*Net web-based system in May 2006 found that one of the core behaviors performed by municipal firefighters was to “participate in courses, seminars and conferences, and study fire science literature, in order to learn firefighting techniques,” which had a task importance rating of 79 out of 100. The work activity of “updating and using relevant knowledge,” which requires the ability to read and comprehend written materials, was rated 71 on the work-activity importance scale. Finally, reading comprehension was rated 76 on the skill importance scale. Based on this O*Net data, it appeared clear that the ability to read, comprehend, and later recall learned information was required in order to be a successful municipal firefighter.

Another source of information that is commonly used to identify important work behaviors and the knowledge, skill, and ability to perform those behaviors is to directly query those who hold the target job position and/or their direct supervisors. Job experts consisting of firefighters and their direct supervisors from the two fire departments involved in the current study participated in a content-related validation study designed to address the requirements of the federal *Uniform Guidelines on Employee Selection*
Procedures (Uniform Guidelines, 1978). These job experts indicated that it was critical that firefighters be able to read and learn printed materials both during the training academy and after the academy was completed. More specifically, these job experts indicated that their cadets typically spent three hours reading each day in academy classes and about five hours reading during each day when not in academy classes (Firefighter Selection, Inc., FSI, 2003). The job experts also unanimously indicated that they themselves could not have succeeded in the entry-level fire academy without the ability to read, comprehend, retain, and later recall written material, while eleven of the twelve job experts indicated they could not have adequately performed the duties of a firefighter without these abilities (FSI, 2003).

Similarly, in 1993, job experts from a large metropolitan fire department in the western United States indicated that firefighter cadets spent an average of five hours reading during each “active” day at the fire academy and an average of four hours of reading per day for each “off” day (Biddle, 1996). In addition, the job experts from the 1993 study indicated that over 3,000 pages of written material must be read, studied, and learned by newly hired employees in the six to ten weeks during which they attend the academy (Biddle & Associates, 1997). Finally, job experts from dozens of professional and volunteer fire departments from across the United States have used a content-related strategy to validate the use of previous versions of the entry-level firefighter TPM test.

In addition, the courts have accepted the use of previous editions of the firefighter TPM test as being valid whenever it was challenged. For example, the 1982 case, Martinez v. City of Salinas, found that the second edition of the firefighter-TPM test met both Title VII and Guidelines criteria for job-relatedness. Findings from this case also
revealed that the TPM test had lower levels of adverse impact against Hispanics when compared to traditional cognitive ability tests. In 1996, the third edition of the firefighter TPM test was also found to be job related and appropriate for entry-level selection purposes in *United States v. the City of Torrance*. These court decisions lend support to the assertion that the TPM testing can be a valid approach to making selection-related decisions when properly constructed.

*Criterion-related validity of trainability testing and of previous versions of entry-level firefighter TPM testing.* Reilly and Israelski (1988) conducted a meta-analytic study that examined 51 trainability coefficients from twelve different studies. The uncorrected validity coefficient they reported for job performance was $r = 0.39$ ($n = 800$) and the uncorrected validity coefficient for training performance was $r = 0.49$ ($n = 2,089$). Notably, Reilly and Israelski (1988) found that the validity evidence for job performance criteria was more generalizable than for training criteria.

Similarly, a meta-analysis by Robertson and Downs (1989) resulted in observed validities for trainability tests of $r = .41$ for ratings, $r = .48$ for errors in training performance, and $r = .24$ for predicting job performance. One possible explanation for these findings is that trainability testing has a greater correlation to what occurs during training than to what occurs during job performance.

During their examination of six trainability tests, Reilly and Isarelski (1988) found considerable variability in the validity of trainability testing across various criteria. On the low side, they found that the telephone facilities assigners’ trainability test had a validity of $r = 0.39$ ($n = 77$) for job performance ratings. On the high end, they found that
the Berger Aptitude for Programming Test (BAPT; Berger, 1987) had a validity of $r = 0.72$ ($n = 44$) for performance during computer programmer training.

While findings from the research cited above indicate that trainability testing can be very beneficial for predicting training and work performance, one limitation of trainability testing, as noted by Robertson and Kandola (1982), is that the predictive validity of trainability tests may attenuate over time. Robertson and Downs (1989) performed a meta-analyses of trainability test data using two different predictor-criterion pairs (total $N_1 = 2,542$ and $N_2 = 2,772$), which led them to suggest that trainability work-sample tests were more accurate at predicting short-term training success than longer-term training success. However, Muchinsky (1986) pointed out that predictors other than trainability tests had also shown similar attenuation issues and yet were still very useful for pre-employment selection purposes. While attenuation over time may be of concern, it does not appear to be more of a concern for trainability tests than for many other types of selection devices that are used to aid in the prediction of success on a job. As with many other types of tests, users should take potential attenuation of the test score-performance relationship into account when using trainability testing as part of an overall employment-selection process.

Since no criterion-related studies of the TPM test version examined in the current study had been conducted, a review of the empirical relationship between performance on two previously-used TPM tests and firefighter performance in both the academy and on the job is presented. A 1982 study found an uncorrected correlation of 0.54 between TPM test scores and fire academy mid-term grades for all members of two fire academy classes in a southern California city ($p < .01$, two-tail, $n = 14$ and $n = 16$; Biddle & Associates,
1997). In addition, the smaller of the two academy classes revealed a significant uncorrected correlation of 0.717 between TPM test scores and final overall Fire Academy scores ($p < .01$, two-tail, $n = 12$). Overall academy scores for this sample consisted of a combination of mid-term and final academy test scores, an equipment manipulation score, note-taking score, and attendance. Also, in 1996, the uncorrected correlation between TPM test scores and first-year firefighter performance appraisal scores for a fire department in the western United States was found to be 0.387 ($p < .05$, $n = 40$) (Biddle, personal communication, December 9, 2004).

The limited criterion-related data presented here suggests that TPM testing has a relatively strong relationship to the firefighter job and academy performance. Of course, additional criterion-related studies would be needed in order to determine the transportability of this type of testing across employers and geographic locations.

*Construct validity of TPM testing in general.* In addition to the content- and criterion-related validity evidence presented above, there is strong theoretical rationale for focusing on the attainment of declarative knowledge during pre-employment testing for entry-level firefighter job applicants. For example, a learning model put forth by Kraiger, Ford, and Salas (1993) postulated that during initial learning acquisition phases, learners focus on declarative knowledge. Kraiger et al. further indicated that most theories of cognitive skill development denoted the acquisition of declarative knowledge as preceding higher-order skill development and that it is appropriate to specify and measure trainees’ retention of declarative knowledge in the initial stages of training. Kanfer and Ackerman (1989) also indicated that declarative knowledge is a precursor to procedural knowledge and skills.
Along similar lines, Campbell, McCloy, Oppler, and Sager (1993) created a general model of performance which postulated that performance is a joint function of declarative knowledge, procedural knowledge and skills, and motivation. The key to all of these models is that the attainment of declarative knowledge is fundamental to successful performance. Thus, a test that measures the ability to learn and recall declarative knowledge, such as the TPM test, would appear to be built on a strong theoretical foundation.

**Relationship of Trainability Tests and General Mental Ability**

The question was raised by Roberson and Downs (1989) as to whether trainability testing was merely measuring cognitive ability and whether it could provide useful prediction beyond that of cognitive ability alone. Along similar lines, Lubinski (2000) asserted that we should discard the view that mental tests and learning tasks measure distinctly different abilities.

Trainability tests do tend to be correlated with scores on measures of cognitive ability. For example, Roth, Bobko, and McFarland (2005) reported a general correlation of \( r = 0.32 \) between work sample tests (if we consider trainability tests to be a type of work sample test) and measures of cognitive ability, which they said was likely downwardly biased by range restriction. To address this concern, they separately examined the validity of a number of military studies that were corrected for range restriction. These military studies suggested that the true relationship between work-sample tests and cognitive ability could be 0.48 (Roth et al, 2005). Similarly, Reilly and Isarelski (1988) found that two types of programming trainability tests were significantly
correlated with a measure of cognitive ability \( r = 0.57, n = 44 \) and \( r = 0.60, n = 68 \) respectively).

However, studies have also suggested that trainability testing provides significant incremental validity over cognitive ability. For example, both of the tests studied by Reilly and Isarelski added significant incremental validity for the prediction of programming proficiency beyond that predicted by the measure of cognitive ability alone. Furthermore, since trainability tests can be considered a subset of a type of work sample testing, it would appear appropriate to point out that Schmidt and Hunter (1998) estimated that there would be a 24% gain in the prediction of job performance if work sample testing were used to supplement the testing of cognitive ability.

*Trainability Testing as an Alternative Selection Methodology to Reduce Adverse Impact Associated with Traditional Measures of Cognitive Ability*

As pointed out in Chapter 1, a basic assumption of this study was that a measure of learning (i.e., as operationalized by trainability testing or the TPM) would have lower adverse impact than a measure of cognitive ability. Surprisingly, there has been less published research on this issue than one would think.

Researchers consistently have reported that adverse impact is lower against minorities than Whites when trainability tests have been used in the past (e.g., Reilly & Manese, 1979; Iasacson & Brown, 1993; Higuera & Riera, 2004). Unfortunately, published standardized-difference subgroup data for job applicants on work sample tests appears to be virtually non-existent. For example, Roth, Bobko, and McFarland (2005) reported that they were unsuccessful in finding studies containing standardized subgroup
differences on work sample tests for job applicants during their search of PsychINFO, General Business File, Ingenta, and Expanded Academic ASAP electronic data bases.

One unpublished study conducted by Harris and MacLane (1988) compared the performance on paper-and-pencil work samples and cognitive ability tests given by the federal government. They found the Black-White subgroup differences on the work sample test were one standard deviation, whereas the subgroup difference on the cognitive ability test was about one and three-quarter standard deviations. Even though the subgroup difference on the work sample test was relatively high, it was still 43% less than the difference on the cognitive ability test. Also, when they conducted a test bias analysis, they found that there was no indication of test bias for two of the positions for which applicants were tested, and bias in favor of Blacks for a third position. Furthermore, they reported that the work sample test had equal or greater validity than the cognitive ability test.

The standardized Black-White mean differences when previous versions of TPM-style tests were used for firefighter selection purposes has typically ranged from .30 to .60 (Stacy Bell, personal communication, April 12, 2006). This is much lower than the generally-accepted Black-White mean difference of one standard deviation that is typically reported in the literature (Roth, BeVier, Bobko, Switzer, & Tyler, 2001). It is also lower than the .72 Black-White mean difference in ability levels that Roth et al. indicate one would expect for applicants who pursue the moderately-complex job of being a firefighter.

An examination of the data from a study by Benver (1996) of TPM testing used in 1994 and 1995 by a city in Kentucky for the selection of entry-level police officers found
a Black-White standardized mean difference of $d = .734$ (Black $n = 126$, White $n = 536$), which was somewhat higher than was reported by Bell. However, as Benver noted, race in the city where the testing he examined took place was “a general indicator of SES (socio-economic status) – with all of the educational and cultural characteristics associated with SES” (p. 141). He further indicated that this likely resulted in much lower reading and comprehension abilities for Black job seekers than for White job seekers. Even so, the subgroup mean difference in his study was still lower than the generally accepted difference found on cognitively-loaded measures.

Furthermore, it has been reported that TPM type testing results in less adverse impact than some other types of written tests, when adverse impact is operationalized in terms of the 80% rule. Campbell (1982) examined several fire departments that were in relatively close geographical proximity to one another in Northern California. He reported that there was no violation of the 80% adverse-impact rule for any of the five fire departments that used TPM testing for selecting their firefighters. Conversely, there were 80% rule violations for all of the other ten fire departments he examined that used what he referred to as entry-level firefighter “aptitude tests” for selecting entry-level employees. Of course, this study could be criticized in that Campbell was comparing different fire departments that may have had varying applicant populations, cutoff scores, and selection ratios. The use of the 80% rule as an indicator of adverse impact is very influenced by situational variables and, by design, Campbell performed the study across jurisdictions.

Even so, there appears to have been some hesitancy on the part of employers to use trainability tests for selecting firefighters. For example, Hunter and Schmidt (1996)
presented several reasons why measures of work performance using work-sample tests, other than physical ability tests, have not historically been used for firefighters, including: the large number of job tasks performed by these employees, the difficulty of staging accurate work under stress or in the presence of physical danger, and the extreme cost of doing so. However, they appear to be referring only to physical tasks that firefighters perform and not to other types of job-related tasks such as reading and learning new information which are commonly measured using paper-and-pencil tests. Paper-and-pencil tests typically present little physical danger and are often relatively inexpensive when compared to physical-ability testing (Landy, 1996).

In summary, TPM testing appeared to be a job-related and valid alternative to using measures of pure cognitive ability for firefighter selection purposes. On the basis of the limited findings of relatively low levels of adverse impact reported when trainability tests have been used in general and, more specifically, when TPM tests have been used during previous employment-selection testing situations, TPM tests would be expected to have lower adverse impact than cognitive ability tests. As previously indicated, it was an assumption of this study that the TPM would have less adverse impact than a cognitive ability test. However, given the limited literature and research on this question, it will be confirmed as a part of testing the first hypothesis:

*Hypothesis 1.* The Black-White subgroup mean difference on the measure of learning on the TPM test will be smaller or less than the subgroup mean difference on cognitive ability.
Potential Alternative Explanations for Lower Subgroup Differences During TPM Testing Compared to Cognitive Ability Testing

In Chapter 1, two models were developed for the purpose of predicting performance on TPM tests. It was thought that a better understanding of performance on TPM tests would reveal why TPM tests would reduce adverse impact. However, if the TPM resulted in lower levels of adverse impact primarily because it was capturing high levels of random error, then it would be unlikely that the TPM would be a valid predictor of firefighter job performance. In addition, in order to consider TPM testing as an alternative to cognitive ability testing, we would need to have some evidence that it is a valid predictor of job performance, or in this case firefighter performance. It was found, and previously indicated in this chapter, that the TPM appears to possess adequate psychometric properties. That is, the TPM possesses adequate reliability, content validity, criterion-related validity, and construct validity. Therefore, it does not appear likely that the reduction in adverse impact is due to random error.

It seemed possible that factors other than cognitive ability, such as the amount of time a person studied the TPM manual and motivational beliefs, might have contributed to the prediction of TPM scores in such a way that this natural interaction resulted in reduced subgroup differences on the TPM test. Asher and Sciarrino (1974) posited a work-sample interaction hypothesis, such that performing complex work-related tasks can elicit interactions among aptitudes rather than simply having an additive effect. Thus, they felt, work sample testing would likely result in a better prediction of job performance than additively combining measures of single traits or aptitudes. This would appear to allow the amount of time the TPM manual was studied and motivational
beliefs, such as motivation to study and study self-efficacy, to influence TPM test performance in a job-related fashion above and beyond cognitive ability.

Referring back to Figure 1 in Chapter 1, according to the Simple Model, there should be a “natural” interaction between cognitive ability and time spent studying. If there are no race differences in time spent studying, and if time spent studying is correlated with TPM performance, then this should result in a TPM score that has a smaller race difference than does the cognitive ability test score. However, cognitive ability will remain an important predictor of the TPM score.

Role of Cognitive Ability and TPM Test Performance

Cognitive ability is necessary for learning and job performance (Gottfredson, 2002; Palumbo, Miller, Shalin, Steele-Johnson, 2005; Schmidt & Hunter, 2004). Schmidt and Hunter (2004) asserted that people who possess greater levels of cognitive ability generally learn more job knowledge quickly and also acquire more of that knowledge, which results in higher levels of job performance. More specifically, Schmidt and Hunter (1998) indicated that measures of cognitive ability have been shown to be one of the best and most popular predictors of learning. Furthermore, a number of researchers (Hunter and Hunter, 1984; Schmidt, Hunter and Outerbridge, 1986; Schmidt and Hunter, 1992) have reported that cognitive ability is strongly correlated with job-related learning.

Jensen (1998) reported that cognitive ability has been shown to be predictive of training performance in a wide variety of situations including school, formal job training, and on-the-job training. Further, Colquitt, LePine, and Note (2000) conducted a meta-analytically based path analysis that showed that cognitive ability directly influenced the attainment of declarative knowledge. Since these and many other studies (e.g., Hunter &
Hunter, 1984; Besetsny, Earles, & Ree, 1993; Ree & Earls, 1991; Roth & Campion, 1992; Roznowski, Dickter, Hong, Sawin, & Shute, 2000; Thorndike, 1986) have consistently linked cognitive ability to learning throughout the literature, the following prediction was made:

Hypothesis 2. Cognitive ability will be positively related to learning as measured by the TPM test.

Study Time and Performance

Previous examinations as to whether study time plays an important role in performance have resulted in inconsistent findings (Rau & Durand, 2000). For example, some researchers (e.g., Blustein, Judd, Krom, Viniar, Padilla, Wedemeyer, & Williams, 1986; Cheung & Kwok, 1998; Michaels & Miethe, 1989) reported a positive correlation between study time and school grades. Schuman, Walsh, Olson, and Etheridge (1985) found a nonsignificant positive relationship between study time and grades. Conversely, Greenwald and Gillmore (1997) and Olivares (2002) found inverse relationships between study time and grades with \( r = -.15 \) and \( r = -.19 \) respectively. Furthermore, Olivares found that grade inflation, school course difficulty, and student cognitive ability all moderated the time-grade association, and even postulated that any relationship between study time and grades may be spurious.

Peverly, Brobst, Graham, and Shaw (2003) examined an intervention designed to improve college students’ self-regulation of allowing for extra study time prior to taking a test and found that there was no relationship between the amount of time studied and students’ prediction of performance. Finally, Walberg (1988) noted that the fraction of lesson and study time that students actually spend on learning activities is more important
than simply time spent in school. This last finding is extremely important as some studies appear to have used time spent in school as a proxy for time spent studying or preparing, which might have led to erroneous assumptions about the potential relationship of study time and TPM-style test performance.

When attempting to examine the relationship between school grades and other criteria, the issue of grade inflation has been of specific concern to researchers. Jennings (1995a, 1995b) indicated that teachers felt pressured to grade more leniently than they had in the past. Sternberg, Grigorenko, and Bundy (2001) pointed out that this grade inflation has made it much more difficult to obtain significant correlations between school grades and other criteria. This is especially true since, as Sternberg et al. pointed out, corrections for range restriction cannot be made unless one knows the parameters of the true grade distribution. The severity of this problem was decidedly illustrated in findings from a government survey, whereby results indicated that a math grade of “A” in high poverty schools closely resembled a math grade of “D” for students in the most affluent schools (Office of Educational Research and Improvement, OERI, 1994).

Some of the inconsistent research findings might also be due to both the type and structure of the material being studied and the study time allotted. A widely cited article by Son and Metcalfe (2000) examined 19 published reports that contained 46 treatment conditions concerning metacognitive control in study-time allocation. However, none of the examined studies presented textual material in narrative form. Instead, the vast majority presented single words or very brief word combinations to participants for extremely brief periods of time. For example, they examined a Mazzoni and Cornoldi (1993) study that reported “study time” was inversely correlated with the probability of
recalling items, which Mazzoni and Cornoldi had referred to as a “labor in vain” effect. It is important to note that the Mazzoni and Cornoldi study presented single sentences consisting of two nouns and a verb for study-time periods of between 2.5 and 7.5 seconds. Nelson and Leonesio (1988) also reported a labor in vain effect based on a study where they presented trigrams to participants though, in the condition that emphasized learning accuracy, the average time studying each trigram was only 5.4 seconds.

Quite the opposite occurred when Son and Metcalfe presented eight six-page biographies to participants. During that part of their study, test performance increased as study time increased. So, it would appear unwise to categorically infer a labor in vain effect (based the parameters of research studies that are measuring the learning of very short bits of information) for a narrative text that is studied for much longer periods of time.

Prior empirical evidence specifically examining the relationship of the TPM manual usage to TPM test scores appears to be limited to two sources. First, Biddle found that the City of Louisville’s 1992 use of a TPM-testing process for the selection of law enforcement officers resulted in an uncorrected correlation between TPM test scores and self-reported TPM-manual study time of $r = .22$ for White applicants and $r = .35$ for Black applicants ($n = 771$) (D. A. Biddle, personal communication, October 20, 2003). However, Biddle’s findings should be used cautiously since the data from the 1992 TPM-testing process is no longer available for further examination or analysis.

The second source of information concerning the relationship between the use of a TPM manual and TPM test scores was from Benver (1996). Benver examined a TPM test that was used in Louisville for entry-level police officer selection in 1994 and 1995.
Interestingly, this is the same TPM manual and test that was used in the 1992 Louisville police-selection process examined by Biddle. Benver reported a positive relationship between use of the TPM manual and TPM test scores of $r = .304 \ (n = 555)$. In order to create a measure for that predictor, Benver combined responses from survey items tapping information concerning the number of weeks that test takers indicated they “used” the manual, number of times per week they “used” the manual, and applicant study methods to create the predictor for that calculation (p. 110). Unfortunately, the cognitive ability of the test takers was not recorded or reported by either Biddle or Benver.

In summary, inconsistent results have been reported by researchers concerning the relationship between study time and learning performance. However, these inconsistent findings could have resulted from some of the studies inappropriately combining short-term and longer-term memory measures. In addition, when the relationship between study time and school grades was examined, other potentially contributing factors, such as grade inflation or school course difficulty, appeared to have been overlooked at times.

It does appear, however, that studies that examined peoples' ability to read and learn text presented in paragraph form have reported significant relationships between the amount of time that the material was studied and the accuracy of later recall. More specifically in regard to the current study, research studies that had specifically examined the relationship between the use of a TPM manual and performance on the TPM test reported finding significant, positive relationships, all of which led to the following prediction:
Hypothesis 3a. The self-reported amount of time spent studying the TPM manual will be positively related to TPM test scores.

Potential Moderator Effect of Cognitive Ability on Study Time

The Interactive, Compensatory Model of Learning (ICML; Schraw, Brooks, & Crippen, 2005) and the Compensatory-Encoding Model (C-EM; Walczyk, 1993, 2000) both posit that weakness in one area, such as poor reading skill or lower cognitive ability, could be compensated for through the use of other effective learning strategies. Strategies are learning tactics that are intentionally used to accomplish a specific purpose or goal (Dole, Duffy, Roehler, & Pearson, 1991). In the ICML model (Schraw & Brooks, 2000), strategies serve a two-fold purpose: they enable those with lower cognitive ability to more efficiently and systematically approach problems and they can increase positive motivational beliefs.

While the overall ICML model has not been tested directly, many of the assumptions on which it is based have been tested (G. Schraw, personal communication, December 27, 2004). For example, Markwell (2004) found that learning was enhanced through the use of tactics, such as budgeting time to study and forming study groups, which he claimed demonstrated the manner in which the ICML model functions in the real world.

Some of the behaviors that help make up the ICML model, especially those dealing with learning strategies, would appear to allow those with lower ability levels to compensate better under a mild time-pressure condition than under a more restrictive time condition, since a mild time-pressure condition would allow more time for a person to read and study the material to be learned. This is consistent with the main thrust of the
current study, which hypothesized that learners with reduced ability levels were more likely to compensate better when longer study time is allowed.

A second compensatory-learning theory of interest to the current study was the Compensatory-Encoding Model (C-EM; Walczyk, 1993, 1995, 2000). The C-EM model specifically posited that allowing additional time for reading and studying provides the opportunity for low ability learners to use a greater number of learning strategies and/or more effort to compensate for their lower ability level. Walczyk and colleagues (Walczyk, 1993, 2000; Walczyk & Griffith-Ross, 2007; Walczyk & Hall, 1989; Walczyk, Kelly, Meche, & Braud, 1999; Walczyk, Marsiglia, Bryan, K. S., & Naquin, 2001; Walczyk, & Taylor, 1996) have reported extensive support for this model. Furthermore, Walczyk and Griffith-Ross (2007) reported that this model has been demonstrated both with children in grade school (National Reading Panel, 2000) and with adults (Walczyk, Kelly, Meche, & Braud, 1999; Walczyk & Taylor, 2000; Walczyk, Marsiglia, Bryan, & Naquin, 2001).

This interaction might be partially due to the struggling that lower-ability readers endure when they are reading and studying, since Salomon and Globerson (1987) asserted that some struggle with reading has been shown to help keep learners engaged. Walczyk, Kelly, Meche, and Braud (1999) found that reading under mild time pressure increased a reader’s mindfulness more so than reading under strong- or no-time pressure conditions. Along similar lines, Walczyk and Griffith-Ross (2007) advised that some highly-fluent readers might not be engaged by the text they are reading, which could cause their minds to wander. Walczyk and Griffith-Ross postulated that because of this,
high-ability learners would be unlikely to use additional learning strategies or effort while reading, even when provided extra time to read or study.

It is interesting to note that research in other areas has found a similar interaction effect for cognitive ability. For example, a study by Frese, Grabarkiewicz, van Steekelenburg, and Escher (2002) found that cognitive ability moderated the relationship between planning and business performance, such that those with lower levels of cognitive ability benefited more from increased levels of planning than those with higher levels of cognitive ability. In other words, those with lower levels of cognitive ability appeared to be compensating through additional planning.

However, additional time studying by those with lower levels of ability was likely to only take place if those with low ability levels recognized that they were not learning proficiently. Walczyk and Griffith-Ross (2007) indicated that struggling readers may not recognize their reading-related shortfalls, and thus not recognize when it would be appropriate to use compensatory learning strategies. This ability is often referred to as metacognitive skill (Metcalf & Shimamura, 1994) and is primarily volitional or conscious (Shimamura, 2000). Discussion of conscious metacognition in psychology (or introspection as it was known then) dates back to the 19th century (e.g., James, 1890). However, metacognition can also be an unconscious process (see Reder & Schunn, 1996 for a review of unconscious metacognition).

Based on these studies, the following hypothesis was proffered:

Hypothesis 3b. The effect of self-reported study time on TPM test scores will be moderated by cognitive ability, such that study time will result in higher TPM scores for those lower in cognitive ability as compared to those higher in cognitive ability.
**Expanded Model**

It appeared likely that individual difference variables would influence the amount of time the job candidates studied the TPM manual. A graphic model of the factors proposed to be related to time studying the TPM manual was presented in Chapter 1, Figure 2. Basically, two variables were identified that might be related to time spent studying: motivation and self-efficacy.

**Effect of Motivation on Study Time and Test Performance**

Olilvares (2002) indicated that study time is a behavior that can be inferred from motivation – the stronger a person is motivated, the longer they will study. To examine that relationship in the current study, job applicants were asked to rate how strongly they agreed or disagreed with the statement “I felt extremely motivated to study the Test Preparation Manual (TPM) to prepare for today’s test.”

In general, motivation theory is based on the work of Vroom (1964), which led to the conceptualization of performance as a function of ability and motivation ($P = F[M \times A]$). This conceptualization has been supported in the literature (e.g., Hollenbeck, Brief, Whitener, & Pauli, 1988). Gottfredson (2002) further refined this model by indicating that performance is a combination of “can do” (i.e., ability), “will do” (i.e., motivation), and “have done” (i.e., experience) factors (p. 37; also see Gottfredson, 2003 for additional discussion of these factors).

As Gottfredson (2002) pointed out, the “will do” factor correlates very little or not at all with cognitive ability. With this in mind, it would appear to be helpful for organizations to determine whether a firefighter applicant possesses the “will do” propensity to succeed in the fire academy and on the job in addition to measuring $g$. As
indicated earlier, both cognitive ability and past behavior have been touted as being the best predictors of future performance. Thus, it would seem logical that a cognitively-loaded testing process in which motivation to study plays a key role for success, such as the TPM test, would be ideally suited to help predict future job and training performance with regard to both “can do” and “will do.”

Interestingly, however, Ryan, Ployhart, and Friedel (1998) found motivation to be unrelated to attendance at firefighter applicant test preparation sessions. They also found that Black job applicants reported being less motivated than White applicants. The authors conceded that one serious limitation of their study was that motivation was measured after the test was given which may have led to rationalization on the part of the respondents. Finally, it should be noted that attendance at the preparation sessions during the Ryan et al. study was unrelated to overall performance in the ability test. While this may seem counter-intuitive, it should be noted that the ability test examined during Ryan et al.’s study (which measured reading comprehension, spelling, and the ability to answer questions about passages heard aloud) did not appear to be the type of exam that would have been influenced by the relatively short-term preparation sessions.

The TPM test, however, does appear to be the type of exam during which those who prepare more would perform better. Thus, in keeping with the generally-accepted concept that performance is a function of ability and motivation, it appeared very likely that those who indicated they were more motivated to study the TPM manual would have studied it longer than those who felt less motivated, resulting in the following prediction.

A logical inference could be made that motivation to study the TPM manual could serve as a proxy for motivation to perform well on the TPM test, especially since
applicants were informed at the time they applied for the position that those who studied the TPM manual longer typically performed better on the TPM test. Previous researchers (e.g., Arvey, Stickland, Drauden, & Martin, 1990; Sackett, Schmitt, Kabin, and Ellingson, 2001) have indicated that a person’s motivation has the potential to influence test performance. Furthermore, even though Ryan, Ployhart, and Friedel (1998) reported that motivation was unrelated to attendance at the preparation sessions, they did find it to be positively related to test performance.

Thus, motivation to study should be related to both time spent studying and TPM test performance. It appeared unlikely that motivation to study the TPM manual would be the direct cause of better performance on the TPM test. Instead, it seemed much more likely that this effect would be primarily explained through the studying of the TPM manual.

Based on this previous review of the likely relationship between motivation, study time, and TPM performance, the following hypotheses were proposed:

*Hypothesis 4a.* Self-reported levels of motivation to study the TPM manual will be positively related to the self-reported amount of time spent studying.

*Hypothesis 4b.* Self-reported levels of motivation to study will be positively related to TPM test scores.

*Hypothesis 4c.* The effect of motivation to study on TPM test scores will be mediated by study time.

**Self-efficacy, Study Time, and TPM Performance**

A core self-evaluation trait that has been found to be related to performance is self-efficacy (Judge & Bono, 2001). Self-efficacy refers to the subjective probability that
a person is capable of successfully executing a task (Bandura, 1977, 1982). Bandura (1989) postulated that beliefs of self-efficacy can enhance or impair performance through its effects on cognitive, motivational, or affective intervening processes. To this end, people with higher levels of perceived self-efficacy are more likely to persist longer, use more regulatory and deep-level strategies, and achieve higher grades than those who are unsure of their ability to perform (Bandura, 1997; Gist & Mitchell, 1992; Pajares, 1996). Bandura (1984) went so far as to broadly assert that a person’s belief in their personal competence touches almost everything they do.

With regard to the current study, Schunk (2003) indicated that an individual’s self-efficacy promotes learning, with Zimmerman (1998) indicating that the correlation between study-strategy use and self-efficacy beliefs in his study ranged from .40 to .46. Other researchers have also reported a significant positive relationship between self-efficacy and performance. For example, Hattie and Purdie (1999) found a correlation of \( r = .215 \) (\( p < .05 \)) between self-efficacy and academic performance. Similarly, a study by Ryan et al. (1998) of firefighter job applicants found test self-efficacy to be significantly related to ability-test performance.

Pajares (1997) indicated that over 20 years of research has supported Bandura's (1986) proposition that self-efficacy beliefs mediate the effect of skill and ability on performance. This is done by influencing a person's efforts and persistence (Bouffard-Bouchard, 1990; Schunk & Hanson, 1985). In other words, the more one believes he or she is capable, the more likely they are to work harder and longer to achieve a goal. Thus, it appeared that those with higher levels of study self-efficacy would be more likely to generate a greater effort to study longer.
Bandura (1986) felt that self-efficacy beliefs were likely to be more related to performance when the self-efficacy beliefs closely corresponded to the task being assessed. Research, especially in the area of learning, appears to have borne this out. For example, Multon, Brown, and Lent (1991) conducted a meta-analysis of 36 studies on the relationship between self-efficacy and academic performance. They found that self-efficacy was more related to students’ performance in course work ($r_u = .36$) than to their performance on standardized tests ($r_u = .13$). In other words, self-efficacy was more strongly related to performance on measures of coursework, which was an outcome over which students had potentially greater control by increasing their effort and persistence when preparing, than for standardized tests, which are likely to be less influenced by preparatory effort and persistence.

In the current study, self-efficacy was measured by a single item that asked the participants to indicate their level of belief as to whether the more they studied the TPM manual the better they would perform on the TPM test. This item seemed in line with the theory of perceived behavioral control (Ajzen, 1985; Ajzen & Madden, 1986), which posits that people will attempt to perform behaviors to the extent that they have confidence in their ability. The participants in the current study had been informed in advance that the test closely mimics reading and learning activities of firefighters in both the academy and on the job. It therefore seemed likely that, under the theory of perceived behavioral control, applicants who believed that the more they studied the higher they would score on a test would be more likely to study longer for that test. Furthermore, it also appeared likely that those who studied longer would also perform better on the test.
The current study predicted that self-efficacy beliefs would be positively related to both studying and TPM test performance. Job applicants who possessed stronger study self-efficacy beliefs would be more likely to study longer and more diligently and therefore perform better on the TPM test, than those with lower self-efficacy beliefs. As mentioned in the introduction, previous research suggests that higher levels of study self-efficacy, in and of itself, were likely not a direct cause of better performance on the TPM test. Instead, it appeared that a relationship between study self-efficacy and TPM test performance would take place primarily through the studying of the TPM manual. Thus, the following hypotheses were offered:

*Hypothesis 5a.* Self-reported levels of study self-efficacy will be positively related to the self-reported amount of time spent studying.

*Hypothesis 5b.* Self-reported study self-efficacy will be positively related to TPM test scores.

*Hypothesis 5c.* The effect of study self-efficacy on TPM test scores will be mediated by study time.

*Mediating Effect of Cognitive Ability on Race Predicting TPM Test Performance*  
Cognitive ability typically serves as the primary explanation for Black-White mean differences when g-loaded measures are used (Jensen, 1985; Gottfredson, 2002; Rushton, 2003). Since the TPM is a strongly g-loaded measure, it was proposed that a relationship between race and performance on the TPM test would likely be explained (i.e., mediated) by differences in cognitive ability, to wit:

*Hypothesis 6.* The effect of race on TPM test scores will be mediated by cognitive ability.
Possible Interaction of Race and Studying

One of the primary goals of the current study was to determine whether Black test takers would benefit more from studying the TPM manual for a longer period of time than White test takers. On its face, the concept that simply extending the time limits for studying text to be tested later would benefit Black test takers over White test takers appeared improbable. This was especially true since researchers have generally indicated that Blacks and Whites appear to benefit equally when active test-taking interventions are offered.

For example, Ryan and colleagues (Ryan, 2001; Ryan, Ployhart, Greguras, & Schmit, 1998) suggested that proactive study interventions, such as test preparation programs or coaching, are likely to benefit both minority and majority members equally. Similarly, Sackett, Schmitt, Ellingson, and Kabin (2001) indicated that the majority of the studies related to the relationship of test preparation programs and test scores revealed they did not appear to reduce the size of subgroup differences. Most interesting in regards to the current study is that Sackett et al. indicated that research does not support the notion that relaxing time limits during testing would help to reduce subgroup differences in test scores. However, it should be pointed out that they did not explore whether relaxed time constraints during the reading and/or learning period prior to testing might reduce subgroup differences.

Benver (1996) found that none of the three study-preparation programs offered to job applicants for entry-level police officers positions helped Black test takers more than White test takers when taking the TPM test. The first program Benver examined consisted of two-hour sessions, during which job candidates could ask questions about
material presented in the TPM manual. The second preparatory program was specifically
designed to teach test-taking skills specifically relevant to TPM-style testing, which
Benver stated was “best described as coaching” (page 33). This second program was
repeated several times during a five-week period and applicants could attend as
frequently as they wished. The third program consisted of a practice test which was
designed to mimic the TPM test used by the city for the selection of their police officers.
In an attempt to reduce potential test anxiety of all of those who attended the practice test
sessions, the practice test was administered in the same room where the actual TPM test
would be later administered.

Based on the reports that preparation interventions have not led to a reduction in
subgroup differences in the past, it was tempting to draw the conclusion that allowing
additional time to study the TPM manual prior to testing would also not likely impact
subgroup performance differences, specifically that Blacks would benefit more from an
intervention than Whites. However, the assertion that subgroup differences are generally
not affected by preparation programs appears to be based on two different fields of
thought. First, some of the studies that have reported no differential benefits for study-
preparation programs focused on preparation for tests of general mental ability, which by
their very nature are not likely to be affected by relatively short intervention programs.
Secondly, other studies, such as those conducted by Benver, focused on a limited number
of the types of strategies that could be used by test takers when preparing for tests.
Further, these preparation programs were not representative of the effort and persistence
associated with the learning of the contents of the TPM manual.
The U.S. Department of Education’s National Adult Literacy Survey (NALS; Kirsch, Jungeblut, Jenkins, & Kolstad, 1993) has reported wide variance in Black and White adult literacy levels on many everyday functions. Literacy, as measured in the NALS survey, is defined as using printed and written information to function in society, achieve one's goals, and develop one's knowledge and potential. According to the survey, two-fifths of the White adults and three-quarters of the Black adults functioned at levels below the range of complex literacy task performance needed for competing successfully in the global economy and fully exercising their rights and responsibilities as citizens (Baldwin, Krisch, Rock, & Yamamoto, 1995). The NALS survey concluded that the impact of education on racial subgroup differences in literacy is "quite substantial" (p. 40).

However, the situation may not be as bleak as it first appears, since providing additional time for reading and studying would likely allow those with lower ability to compensate through the use of additional strategies and/or effort. For example, researchers have found that reading something more than once is often an efficient reprocessing strategy that will aid in improving reading comprehension performance (e.g., Amlund, Kardasz, & Kulhavy, 1986, Barnett & Seefeldt, 1989; Haenggi & Perfetti, 1992). Haenggi and Perfetti found that rereading text was just as effective a strategy for improving reading comprehension as rewriting or rereading notes about the text. Additionally, the Compensatory-Encoding Model of reading (Walczyk, 2000) posits that slower reading rate, looking back, pausing, and other means can all help learners to compensate for small verbal working-memory capacity and less automated reading skills. To that end, a study by Walczyk, Marsiglia, Bryan, and Naquin (2001) found that less-
verbally efficient readers can compensate both behaviorally and strategically when the reading does not occur under time pressure.

Thus, under restrictive time conditions, such as during a traditional standardized reading test, the reading comprehension skill of less-verbally efficient readers will likely be underestimated (Walczyk, 2000). Allowing additional time for reading, especially when appropriate for the job being tested, would appear to be a likely solution to this dilemma.

Since Black learners generally demonstrate lower reading ability levels than White learners (NALS; Kirsch, Jungeblut, Jenkins, & Kolstad, 1993), it appeared possible, based on Walczyk and colleagues’ assertions, that Black job applicants who studied the TPM manual longer would be likely to use more strategies to compensate than Whites who studied a similar amount. Also, as indicated by Salomon and Globerson (1987), some struggle with reading can help to keep readers engaged, whereas Walczyk and Griffith-Ross (2007) reported that some higher ability readers are more likely to become distracted as they study longer.

Research questions are asked in an attempt to fill in gaps that may currently exist in the literature for explaining how or why relationships occur. The following research questions were asked to help clarify whether compensation during reading and learning under extended study-time intervals occurs:

*Research Question 1.* Will time spent studying and race interact to affect learning, as measured by the TPM test, such that Black test takers benefit more from increased study time than White test takers?
Research Question 2. Will the Black-White subgroup mean difference and potential adverse impact on the TPM test decrease as the amount of study time increases?

Study Strategies

This next section examines some of the pertinent literature related to other factors, in addition to the amount of time spent studying, motivation, self-efficacy, and cognitive ability, which might also influence a potential relationship between time spent studying the TPM manual and TPM test performance. Because these remaining sections deal with potential alternative factors that might explain any significant relationship found between study time and TPM test performance, no formal hypotheses are offered. However, following an examination of the literature, a research hypothesis is offered concerning how these additional factors might interact to influence the relationship between the amount of time studying the TPM manual and TPM test performance is offered.

Test preparation training/coaching. There are many different types of test preparation interventions that have been examined in the literature, including coaching. The survey used in this study elicited information concerning potential coaching-related influences by asking if test takers studied in a group and/or used a tutor when preparing for the TPM test. Merrill (1992) indicated that tutors typically provide coaching to students as the students practice the lessons being taught. While studying with a tutor would likely involve a more traditional form of coaching, studying in a group would likely encourage applicants to “coach” one another. For example, those in a study group might offer motivation, test-taking hints, and other strategies to one another, all of which have been identified as types of “coaching” activities in the literature.
Only a very limited amount of research on the relationship between coaching/test preparation and employee selection procedures in particular has been reported (Ryan, Ployhart, Greguras, & Schmit, 1998). For this reason, research from the literature relating to the relationship between coaching and academic performance will be presented first. A review of studies that have more specifically examined the relationship between coaching and employment-selection test performance will follow.

Educational researchers have studied the successful coaching of young adults on measures of cognitive ability for many years. For example, as far back as 1927, Gilmore reported that students could be coached on specific intelligence tests "to the point of increasing their standing and score in intelligence tests even in the case of the material used in coaching being only similar and not identical with that of the basic test" (Gilmore, 1927, p. 321). Linn (2000) pointed out a commonly found pattern by which test scores steadily rise following a new testing program’s introduction and then fall dramatically when a different testing program is introduced. This repeating pattern, referred to as the “saw tooth” phenomenon, is likely due to the test takers being trained or coached on how to respond to the types of questions that each of the testing programs offer (Linn, 2000, p. 7).

One of the biggest problems to overcome when conducting a review of the relationship between test preparation training and performance is, as suggested by Messick (1982), that many of the published studies use different terms to mean the same thing or similar terms to mean different things. Take for instance, Palmer and Buscigilio’s (1996) use of the term “coaching.” These researchers used “coaching” (in reference to aptitude test preparation) to include activities ranging from giving titles of
practice books or study guides to giving the right answers on the tests. Jensen (1980) limited his use of the same term to describe “…instructing the testee in the ‘tricks of the trade’ in test taking, including how to analyze test questions and problems, instructions and demonstrations in working through typical test problems, distributing one’s time most efficiently, and doing many typical practice problems with a tutor or manual providing immediate informative feedback” (p. 591).

There has been a vast amount of research examining the influence of coaching on standardized academically-based aptitude test performance, most of which surrounding the Scholastic Aptitude Test (SAT) and Graduate Records Exam (GRE). The TPM test’s stated goal is to predict fire-academy success where only a high school diploma is required as the academic prerequisite. Therefore, the current review primarily focuses on the SAT-related research as the SAT is typically given at the close of high school as a means of predicting learning performance after high school.

Reynolds, Oberman, and Perlman (1988) found that a 64-hour Preliminary SAT coaching period had significant effects for math scores but not for verbal scores. Brody and Benbow (1990) found that relatively short-term academic training of three weeks did not affect SAT scores, even when the knowledge presented was directly relevant for solving SAT-type problems. Black students from low-income families who were coached for two-and-one half to three hours twice a week for seven weeks in algebraic, geometric, and general mathematics showed significant gains at the .05 level on the SAT (Johnson & Wallace, 1989). Not surprisingly, those with the lowest pre-test scores made the greatest advances.
A meta-analysis of coaching for achievement-test studies by Banger-Drowns, Kulick, and Kulick (1983) concluded that coaching had only a relatively small effect on test performance. However, Crocker and Schmitt (1987) questioned the finding that coaching can only have a weak to moderate effect on test performance. To demonstrate this, they conducted an experiment involving 124 college students, some of whom were coached on strategies for improving multiple-choice item test performance. They found that the test-taking strategy that was taught (i.e., coached) was effective for low-anxious students, neither effective nor ineffective for students who were in the mid-range of test anxiety, and mildly detrimental to high-anxious test takers. They postulated that previously undetected interactions, such as those they reported, could substantially weaken effect sizes or lead to an erroneous conclusion that coaching has no effect. In other words, the weak relationships that have been reported in many studies concerning the relationship between coaching and test performance may be a result of their failure to take into consideration possible interactions between predictors (e.g., student test anxiety levels). In summarizing this issue, Sackett, Schmitt, Ellingson, and Kabin (2001) indicated that the majority of studies on coaching and orientation programs revealed that these programs have little positive impact on the size of subgroup differences.

Another potential issue regarding the relationship between test preparation programs and test performance was suggested by Ryan, Ployhart, Greguras, and Schmit (1998). They postulated that one reason for the finding of a very limited relationship between test preparation programs and test performance may be due to lower-ability performers being more likely to seek out and attend test preparation sessions. Sackett, Burris, and Ryan (1989) noted that the educational literature suggests that the effect size
for the relationship between practice and ability-test scores is smaller for low-ability test takers. However, Ryan et al. (1998) found that attendance by firefighter applicants \((n = 4,025)\) at a preparation program session was unrelated to scores on a combined reading, listening, and spelling test, even when relevant individual differences related to self-selection into the program were controlled. Findings from this study also revealed a significant correlation \((r = .19)\) between study habits and ability test scores, and a significant correlation \((r = .30)\) between motivation and ability test scores (both \(p < .001\)). Interestingly, no significant relationship between attendance at the test preparation program and either anxiety or motivation was found. In addition, lower self-reported motivation levels for Black applicants as compared to White applicants were found and women reported significantly more positive study habits and positive study self-efficacy than men. Unfortunately, it was impossible to compare the study habits and self-efficacy perceptions of preparation session attendees versus non-attendees in the Ryan et al. study as this information was only collected from those who attended the preparation sessions.

Benver (1996) reported mixed results of the relationship between attendance at test-preparation classes and the TPM-test scores of police officer applicants, with a positive relationship between preparatory class attendance and test scores for some of the test administrations, but no relationship for other test administrations. However, even when a positive relationship between attendance at the preparation classes and test performance was reported, the relationship was apparently confounded. Specifically, Benver indicated that those who attended preparation classes also appeared to have studied longer and taken more practice examinations, which could have served as an alternative explanation for their obtaining higher test scores even if they had not attended
the preparation classes. Finally, similar to the results reported by other researchers, Benver found that preparatory programs did not help Black candidates more than they helped White candidates.

Another possible explanation for the potential positive relationship between studying in a group or with a tutor and TPM test scores might be related to the setting of goals. For example, maximum advantage of group study can be obtained only if group members have all read and studied the same materials prior to the study session. This would require that reading and studying goals be set for everyone in the group. Similarly, studying with a tutor would appear to be maximally effective when reading and study goals have been met. This type of goal setting has long been associated with increased performance, as an individual’s goals are immediate regulators of action (Locke & Latham, 1990).

One final note regarding study groups is from the work of Salomon and Globerson (1987) who found that working in study teams fosters students’ mindful engagement of a learning task which can enhance performance. This might also add a further explanation for the potential contribution of studying with a tutor or study group to the prediction of TPM test scores.

Effect of taking practice exams on test performance. Credible feedback about effort frequently improves task persistence (Schunk, 1987; Schunk & Cox, 1986). As mentioned in the previous section, studying in a group or with a tutor is likely to result in task-specific feedback. Another way for learners to receive feedback is through practice questions where the individual can judge whether he or she is successfully mastering the material likely to appear on the test.
In the case of the current study, the employer gave test takers the option to purchase a collection of 50 sample questions to provide a practice opportunity for the test. These practice questions were taken from the same TPM-question item bank as the TPM test questions used during the pre-employment testing process (Stacy Bell, personal communication, January 29, 2004).

Interestingly, research reveals that applied practice is more effective than coaching alone. Specifically, Jensen (1980) and Vernon’s (1960) reviews examining the influence of coaching on ability test scores both revealed that coaching was not effective unless it was accompanied by applied practice. Quinn (1993) found that a practice pre-test led to ability test score improvement by increasing test takers’ awareness of their skill deficiencies. Jackson and McGlinn (2000) found additional support for a practice effect impacting test performance, especially for the higher-performing students.

Other researchers, such as Glover (1989) and Walczyk and Hall (1989), found that taking practice tests that were similar to the final test increased test taker’s accuracy of test performance predictions. A study by Balch (1998) found that students rated practice exams as more helpful in test preparation than only the reviewing of a parallel exam. Balch also found that those who took the practice exam scored higher on the final exam. He postulated that performance on the practice exam may motivate the participants to adjust their subsequent studying but was unable to definitively answer this question due to his study’s design. The current study examines whether those who take a practice exam typically study longer and/or perform better on the test than those who did not.

Maki and Serra (1992) found that multiple-choice practice test questions related to 12 text passages were useful to college students in helping them to predict their overall
performance on a later test of these same passages, but did not enhance their ability to accurately predict later test performance for specific passages. Based on this, they warned against using practice tests for assessing which portion of a text the test taker knows proficiently. However, their suggestion should be taken cautiously since they only provided one practice test question per passage which resulted in extremely low reliability.

Finally, during a study conducted by Ryan, Ployhart, Greguras, and Schmit (1998), applicants for a firefighter position were offered an optional test-preparation session. In that session, the applicants were given practice tests similar to the ability test used during the selection process and were also provided feedback about their practice test performance. A review of ability testing in general and study hints were also offered during the preparation session. However, Ryan et al. found no significant relationship between attendance at the test preparation session and performance on the ability test that was later used as part of the firefighter selection process.

In summary, the role of study techniques on the prediction of test performance as reported in the literature has been equivocal, so there was little theoretical basis on which to form a prediction about techniques that may have been used by job applicants in the current study. However, under the ICML model of learning (Schraw, Brooks, & Crippen, 2005), study techniques can help learners to compensate for weaknesses in other areas, such as low levels of cognitive ability. To examine how various study techniques might help to predict TPM test scores above and beyond the amount of time the TPM manual was studied, the following research hypothesis was tested:
Research Question 3. A combination of study techniques (including studying with a tutor or study group, setting aside certain times to study, and the number of times a practice test was taken) would explain variance beyond that explained by study time and cognitive ability.
CHAPTER III

METHODOLOGY

Participants

Participants were job applicants who were taking a written examination as the first step in the hiring process for entry-level firefighter positions with one of two mid-sized fire departments in the western United States. Applicants were required to pass this written test in order to advance to the next stage in the selection process. Of the 2,195 applicants who took the exam, a total of 2,078 (94.67%) voluntarily responded to a 15-item survey. Only information collected from Black ($n = 157$) and White ($n = 1243$) participants was used in during the analyses conducted during the current study (see Table 1 for demographic breakdown of respondents).

Upon filing a job application, applicants were informed that they must have met the departments’ minimum requirements by the time of the exam, which consisted of a high-school or high-school equivalence diploma and a valid driver’s license. In addition, applicants were informed that they needed to be in possession of a basic Emergency Medical Technician (EMT) certificate by the date of a job offer, if they wished to be hired. It should be noted that EMT certification in the departments’ locale typically required between 17 to 23 weeks of training. It is unknown what percentage of the job applicants were already certified as EMTs prior to their taking the written test.
Table 1

*Race of Test Takers Who Responded to Characteristics Survey*

<table>
<thead>
<tr>
<th>Race</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black/African American</td>
<td>157</td>
<td>7.56%</td>
</tr>
<tr>
<td>White</td>
<td>1243</td>
<td>59.82%</td>
</tr>
<tr>
<td>Hispanic/Latino</td>
<td>363</td>
<td>17.47%</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>113</td>
<td>5.44%</td>
</tr>
<tr>
<td>Native American</td>
<td>41</td>
<td>1.97%</td>
</tr>
<tr>
<td>Other/Did Not Specify</td>
<td>161</td>
<td>7.75%</td>
</tr>
</tbody>
</table>
The archival data used in this study was collected by Firefighter Selection, Inc. (FSI) in their normal course of business as part of their continual review of tests that they develop and employ. Approval for the current study’s protocol was received from The University of Akron’s Institutional Review Board for the Protection of Human Subjects (IRB; S. McWhorter, personal communication, April 5, 2004). It was determined by the IRB that this project represented minimal risk to the participants and was eligible for an exemption from continuing approval under a federal category for the use of data that was publicly available or if the information was recorded by the investigator in such a manner that subjects could not be identified, directly or through identifiers linked to the subjects. A copy of the IRB’s approval letter is attached as Appendix A.

**Demographic Data**

Participants were asked to voluntarily indicate their race, gender, and year of birth in the spaces provided on the test. This data was also collected on the study behavior survey. Table 1 contains race data for all participants who supplied this information. For the Black and White participants, Table 2 provides data on gender, race, age, and highest education level.

**Measure of Study Behavior**

The study behavior survey used in the current study contained 15-multiple choice items with five alternatives for each item to assess test-taker individual-difference characteristics. With the exception of the academic achievement item, all item alternatives were written to approximate an interval scale as linear regression assumes the use of interval data (Size, 1987). Since previous research has indicated that reverse-coded items may produce artifact response factors (e.g., Podsakoff, MacKenzie, Lee, &
Table 2

*Demographics of Black and White Participants Who Responded to Characteristics Survey*

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1285</td>
<td>91.92%</td>
</tr>
<tr>
<td>Female</td>
<td>113</td>
<td>8.08%</td>
</tr>
<tr>
<td>Did not specify</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 40</td>
<td>1318</td>
<td>94.28%</td>
</tr>
<tr>
<td>40 or more</td>
<td>68</td>
<td>4.86%</td>
</tr>
<tr>
<td>Did not specify</td>
<td>12</td>
<td>0.86%</td>
</tr>
<tr>
<td><strong>Highest education level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school diploma</td>
<td>660</td>
<td>47.21%</td>
</tr>
<tr>
<td>G.E.D.</td>
<td>60</td>
<td>4.29%</td>
</tr>
<tr>
<td>Associate Degree/Vocational School Certification</td>
<td>466</td>
<td>33.33%</td>
</tr>
<tr>
<td>Bachelors degree</td>
<td>188</td>
<td>13.45%</td>
</tr>
<tr>
<td>Beyond a Bachelors degree</td>
<td>23</td>
<td>1.65%</td>
</tr>
<tr>
<td>Did not specify</td>
<td>1</td>
<td>0.07%</td>
</tr>
</tbody>
</table>
Podsakoff, 2003), all items were written in a positive direction (see Appendix B for the items and potential responses).

Characteristic areas measured included the amount of time spent studying the Test Preparation Manual (TPM manual), the number of times a practice test was taken, whether the participant studied with a tutor or study group, prior academic achievement, study self-efficacy (the applicant’s belief concerning the extent to which he or she could affect the outcome of the exam results through study), motivation to study, ease of reading the TPM manual, study style/environment, knowledge about how much reading was required of a cadet in the fire academy, and additional items as described in Appendix B. All of the items required the test takers to choose from one of five possible alternatives that were essentially unique to each item.

There are good reasons to be cautious when using self-report questionnaires. However, self-report designs often represent a valid measurement strategy when employed in a sensible design (Howard, 1994; Schmitt, 1994; and Spector, 1994). The following are some of the steps that were taken to minimize potential threats to validity that are sometimes mentioned in association with the use of self-reported data.

First, in keeping with recommendation by Oskamp and Schultz (2003), the survey data-collection forms were visually scanned and data were discarded for those who showed potentially invalid patterns of responding (such as A, B, C, A, B, C). A visual examination was also conducted to determine whether participants were generally responding to items using extreme or midrange response sets (such as consistently choosing items at either end of the scale or showing a central-tendency response bias by consistently choosing items from the middle of the scale). Survey results from two of the
White participants that contained these types of non-valid response patterns were removed from the final data set.

Another area of concern with self-reported information is that participants might attempt to manage the impressions they present by providing exaggerated or false information that would present them in a positive light. It has been suggested that two of the primary triggers for impression management are identifiability (Arkin, Appelman, & Berger, 1980; Bradley, 1978) and consequences, such as social or material outcomes (Schlenker, 1980). These two triggers were addressed in several ways during the current study.

As mentioned previously, participants were informed, both in writing and verbally, that responding to the characteristics survey items was voluntary and that there would be no penalty if they did not chose to respond to the survey. Second, they were instructed that if they chose to respond to the survey they should skip any item that they did not wish to answer or that they were not likely to answer honestly. Finally, they were informed that neither of the two fire departments that were using the test for selection purposes would see their individual responses to the survey items.

Furthermore, the responses to particular survey items were examined to determine whether participants, in general, were attempting to provide a favorable impression to those receiving the information. For example, we would expect participants to indicate that they had studied the maximum number of hours available if they were trying to present the most favorable impression to the two fire departments that were conducting the testing associated with the current study. However, the responses to the five alternatives indicating the number of hours studied during the last two weeks before the
test was taken were almost uniformly distributed (Alternative A – 16%, B – 21%, C - 22%, D - 22% and E – 19%). While this pattern does not discount impression management by some of the respondents, it would indicate that, in general, participants were not attempting to present themselves in only a most favorable light.

Next, responses to survey items were examined to determine if there were any unexpected or unusual response sets. During this part of the examination of the survey data it was noted that 78 of the 2,078 participants (3.77%) indicated they strongly disagreed with the statement that they were smart enough to be a firefighter whereas none of the participants indicated that they disagreed with this statement. The remainder of the participants indicated they neither agreed nor disagreed (1.40%), agreed (11.01%), or strongly agreed (83.82%) with that statement. On the surface, it appeared highly suspect that any of the test takers would honestly believe that he or she was not smart enough to be a firefighter and still voluntarily remain in the pre-employment selection process. Since both random- and deliberately-false responding can attenuate validity, it was decided to exclude the responses of those 78 test takers who indicated they strongly disagreed they were smart enough to be a firefighter from the analyses in the current study.

Pre-employment Selection Measures

The pre-employment selection device examined in the current study consisted of a single, two-part paper-and-pencil test. The first part of the test contained 80 multiple-choice questions, which were designed to measure knowledge of information contained in the TPM manual. The second part of the test contained 28 multiple-choice questions, which were designed to measure cognitive ability or g. All test items were worth one
point. Correct responses from both sections of the test were totaled to create an overall score for the TPM and for cognitive ability.

The version of the TPM manual and the TPM test that were used in the current study are still actively being used for pre-employment selection purposes by other organizations. For this reason, only sample content from a previous version of the TPM manual and sample questions from a previous version of the TPM-style test, which are no longer used for selection purposes, are included in this report (see appendixes C and D). The following is a description of each of the two sections that made up the written test.

**Test preparation manual test (TPM test).** The TPM test used in the current study consisted of 80 multiple-choice items. These items were designed to measure declarative knowledge that was to be learned by studying a 103-page TPM manual. This manual contained information that had been adapted from on-the-job and academy training materials from several fire departments including training materials from one of the two departments involved in the current study. This material was modified, as necessary, so that applicants did not need to possess any specialized training or experience as a firefighter.

Two test items contained in the TPM-style test associated with the current study were removed prior to the calculation of the overall test score because they possessed unacceptable item-level statistical values (S. Bell, personal communication, February 19, 2004). The responses from these two items were not used by the fire departments when making their selection decisions nor were they used when conducting any of the analyses associated with the current study.
The internal-consistency reliability (Cronbach, 1951) of the 78 remaining TPM-style test items was 0.936 \((n = 2,195)\). The U. S. Department of Labor indicates that reliability coefficients of greater than .90 are “excellent” (Testing and Assessment, 1999; p. 3-3).

_Cognitive ability test (CAT; predictor variable)._ The second portion of the written test consisted of 28 multiple-choice items designed to measure cognitive ability. The uncorrected relationship between the scores on these 28 items and job performance from a 2003 concurrent validation study was \(r = .342\) \((p < .000; D. Biddle, personal communication, January 15, 2004)\).

Responses to one of the 28 CAT test items given to the test takers in the current study were not used for selection purposes nor when conducting analyses associated with the current study because of a typographical error that was made when the test was printed. This lowered the total number of CAT items used for employment-selection and analysis purposes in the current study to 27. The internal-consistency reliability (Cronbach, 1951) of responses to the 27 CAT items from the test session associated with the current study was 0.789 \((n = 2,195)\). The United States Department of Labor indicates that reliability coefficients between .70 and .79 are “adequate” (Testing and Assessment, 2000; p. 3-3).

_Design and Procedure_

In the first week of October 2003, firefighter-trainee job applicants were invited by the two participating fire departments to partake in the written selection test scheduled for the fourth week of January 2004. At that time, applicants were provided information about how they could obtain a copy of the TPM manual by mail for $17 (including
shipping and handling) and information about how they could view copies of the TPM manual free of charge at the public libraries in both of the cities that were conducting the testing. Similar announcements were mailed to those who continued to apply for an entry-level position with either fire department up until December 31, 2003, at which time recruitment was closed. Information about the test testing process and TPM manual was also made available on both fire departments’ web sites beginning in early October 2003. All of this information was also available on both fire departments’ web sites until after the written test had been administered.

Prospective test takers were also informed that a practice test, which contained 50 questions similar to those that would be contained on the actual test, was available for purchase only through the mail for seven dollars (including shipping and handling). The TPM manual and the practice test were sent via first class mail to each applicant within one-business day of a written request and payment being received.

Testing was conducted in a large auditorium during four sessions on a single day in January 2004. Each test taker had been randomly assigned by the fire departments’ human resource personnel to one of the four testing sessions and notified in writing in advance as to which session they were scheduled to attend.

At the beginning of each testing session, test-takers were first asked to respond to the 15-item study behavior measure. At that time they were informed, both verbally and in writing, that responding to these items was voluntary and that the information they provided would be used to “study and improve firefighter testing in the future.” They were further told that there would be no penalty if they preferred not to respond to the survey items. After the test takers were given these instructions, they were asked to
indicate their SSNs on the response form so that their responses could later be linked to
their written test scores. Finally, they were told that their names and SSNs on the study
behavior response form would be removed once the linkage was made.

After the survey behavior response forms were collected, the written test and
answer sheet were distributed to the test takers. The test takers were instructed to indicate
their name and SSN on the answer response form. They were also asked to voluntarily
indicate gender, race, and year of birth information if they wished to provide this
information for research purposes only. Test takers were then allowed two-and-one-half
hours to answer the 108-item multiple-choice questions contained on the written test.
Less than one percent of the test takers failed to complete the written test within the
allotted time. Test takers were required to answer all test questions without referring to
any written material or other resources.

In an effort to protect the identity of the participants’ in this study, all information
was collected on scannable-answer forms and then downloaded into a database program
using a machine-scoring device. A trained human-resource analyst then linked
participants’ responses on the characteristics survey to their written test scores and
demographic data using the SSN as the unique identifier. A second analyst independently
double-checked the linkage to ensure accuracy. Once the linkage was completed, specific
identifying information (i.e., name and SSN) was removed to ensure participant
confidentiality prior to my being provided the data. All information resulting from the
analyses is reported in aggregate form to further minimize the potential of identifying
individual test-takers.
CHAPTER IV
RESULTS

In this chapter, the descriptive data are presented and the various hypotheses are subject to statistical testing. In addition, a number of research questions or exploratory hypotheses are investigated.

The descriptive statistics and figures reported in this chapter reflect raw data for easier interpretation, whereas standardized scores or $Z$ scores were used to test the hypotheses. Also, it should be noted that during the analyses, females and males were coded 0 and 1, respectively. Similarly, Black and White participants were coded as 0 and 1, respectively.

Descriptive Statistics, Correlations, and Tests of Demographic Differences in Predictors and Criteria

Means, standard deviations, and correlations of the predictors and criteria are presented in Table 3. Comparisons of the characteristics, demographics, and test-scores were initially examined using independent-sample $t$ tests (see Table 4). Independent-samples $t$ tests were also used to analyze differences between age groups (less than 40 years of age and 40 or more years of age; see Table 5) and by gender (see Table 6). Where Levene’s test for inequality of variance was significant ($p < 0.05$), equal variances were not assumed in the $t$ test.
Table 3

Means, Standard Deviations, and Intercorrelations Between Measures

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2.58</td>
<td>1.94</td>
<td>3.16</td>
<td>3.35</td>
<td>3.60</td>
<td>3.35</td>
<td>3.16</td>
<td>3.04</td>
<td>3.25</td>
<td>3.42</td>
<td>3.04</td>
<td>3.25</td>
</tr>
<tr>
<td>SD</td>
<td>1.44</td>
<td>1.55</td>
<td>1.36</td>
<td>1.35</td>
<td>1.21</td>
<td>1.28</td>
<td>1.32</td>
<td>1.43</td>
<td>1.27</td>
<td>1.49</td>
<td>1.49</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05. **p < .01. ***p < .001.
Table 4

*Characteristic, Demographic, and Test-Score Differences Between Black and White Participants*

<table>
<thead>
<tr>
<th>Measure</th>
<th>White</th>
<th>Black</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>Black</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>Education</td>
<td>1194</td>
<td>2.61</td>
<td>0.82</td>
</tr>
<tr>
<td>High school grade</td>
<td>1192</td>
<td>2.24</td>
<td>0.64</td>
</tr>
<tr>
<td>Tutor or study group</td>
<td>1182</td>
<td>1.81</td>
<td>1.07</td>
</tr>
<tr>
<td>Practice test</td>
<td>1193</td>
<td>2.76</td>
<td>1.43</td>
</tr>
<tr>
<td>Number times TPM read</td>
<td>1192</td>
<td>2.79</td>
<td>1.40</td>
</tr>
<tr>
<td>Easy to read</td>
<td>1155</td>
<td>3.85</td>
<td>0.85</td>
</tr>
<tr>
<td>Motivated to study</td>
<td>1172</td>
<td>3.88</td>
<td>0.96</td>
</tr>
<tr>
<td>Quiet place to study</td>
<td>1166</td>
<td>3.78</td>
<td>1.00</td>
</tr>
<tr>
<td>Total hours studying</td>
<td>1187</td>
<td>2.71</td>
<td>1.21</td>
</tr>
<tr>
<td>Final two weeks studying</td>
<td>1186</td>
<td>3.09</td>
<td>1.31</td>
</tr>
<tr>
<td>Study self-efficacy</td>
<td>1166</td>
<td>4.04</td>
<td>0.94</td>
</tr>
<tr>
<td>Importance of hire</td>
<td>1194</td>
<td>4.65</td>
<td>0.67</td>
</tr>
<tr>
<td>Firefighter self efficacy</td>
<td>1194</td>
<td>4.87</td>
<td>0.36</td>
</tr>
<tr>
<td>Certain time to study</td>
<td>1175</td>
<td>3.73</td>
<td>1.04</td>
</tr>
<tr>
<td>Hours read on job</td>
<td>1135</td>
<td>4.12</td>
<td>0.95</td>
</tr>
<tr>
<td>TPM test scores</td>
<td>1194</td>
<td>53.50</td>
<td>13.54</td>
</tr>
<tr>
<td>Cognitive ability scores</td>
<td>1194</td>
<td>23.33</td>
<td>3.09</td>
</tr>
<tr>
<td>Gender</td>
<td>1194</td>
<td>0.92</td>
<td>0.27</td>
</tr>
<tr>
<td>Age</td>
<td>1186</td>
<td>27.60</td>
<td>5.75</td>
</tr>
</tbody>
</table>

* p < .05. ** p < .01. *** p < .001.
Table 5

*Characteristic, Demographic, and Test-Score Differences between Younger (under 40) and Older (40 and over) Participants*

<table>
<thead>
<tr>
<th>Measure</th>
<th>&lt; 40 years of age</th>
<th></th>
<th></th>
<th>&gt;= 40 years of age</th>
<th></th>
<th></th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>1267</td>
<td>2.61</td>
<td>0.83</td>
<td>66</td>
<td>2.76</td>
<td>0.90</td>
<td>1.44</td>
</tr>
<tr>
<td>High school grade</td>
<td>1266</td>
<td>2.25</td>
<td>0.64</td>
<td>65</td>
<td>2.38</td>
<td>0.55</td>
<td>1.62</td>
</tr>
<tr>
<td>Tutor or study group</td>
<td>1254</td>
<td>1.86</td>
<td>1.12</td>
<td>64</td>
<td>2.16</td>
<td>1.35</td>
<td>1.71</td>
</tr>
<tr>
<td>Practice test</td>
<td>1266</td>
<td>2.75</td>
<td>1.43</td>
<td>66</td>
<td>3.05</td>
<td>1.45</td>
<td>1.63</td>
</tr>
<tr>
<td>Number times TPM read</td>
<td>1265</td>
<td>2.78</td>
<td>1.40</td>
<td>66</td>
<td>3.33</td>
<td>1.33</td>
<td>3.14***</td>
</tr>
<tr>
<td>Easy to read</td>
<td>1224</td>
<td>3.83</td>
<td>0.85</td>
<td>66</td>
<td>3.92</td>
<td>0.93</td>
<td>0.86</td>
</tr>
<tr>
<td>Motivated to study</td>
<td>1241</td>
<td>3.87</td>
<td>0.96</td>
<td>66</td>
<td>4.20</td>
<td>0.83</td>
<td>2.67**</td>
</tr>
<tr>
<td>Quiet place to study</td>
<td>1236</td>
<td>3.80</td>
<td>1.00</td>
<td>66</td>
<td>3.97</td>
<td>0.94</td>
<td>1.46</td>
</tr>
<tr>
<td>Total hours studying</td>
<td>1260</td>
<td>2.73</td>
<td>1.23</td>
<td>66</td>
<td>3.09</td>
<td>1.20</td>
<td>2.33*</td>
</tr>
<tr>
<td>Final two weeks studying</td>
<td>1258</td>
<td>3.10</td>
<td>1.32</td>
<td>66</td>
<td>3.42</td>
<td>1.16</td>
<td>1.93</td>
</tr>
<tr>
<td>Study self-efficacy</td>
<td>1237</td>
<td>4.05</td>
<td>0.92</td>
<td>66</td>
<td>4.11</td>
<td>1.04</td>
<td>0.49</td>
</tr>
<tr>
<td>Importance of hire</td>
<td>1267</td>
<td>4.66</td>
<td>0.64</td>
<td>66</td>
<td>4.56</td>
<td>0.93</td>
<td>-0.90</td>
</tr>
<tr>
<td>Firefighter self efficacy</td>
<td>1267</td>
<td>4.87</td>
<td>0.36</td>
<td>66</td>
<td>4.83</td>
<td>0.41</td>
<td>-0.86</td>
</tr>
<tr>
<td>Certain time to study</td>
<td>1246</td>
<td>3.76</td>
<td>1.03</td>
<td>66</td>
<td>3.89</td>
<td>0.98</td>
<td>1.03</td>
</tr>
<tr>
<td>Hours read on job</td>
<td>1203</td>
<td>4.12</td>
<td>0.95</td>
<td>63</td>
<td>4.05</td>
<td>0.94</td>
<td>-0.59</td>
</tr>
<tr>
<td>TPM test score</td>
<td>1267</td>
<td>52.93</td>
<td>13.80</td>
<td>66</td>
<td>54.88</td>
<td>14.78</td>
<td>1.12</td>
</tr>
<tr>
<td>Cognitive ability score</td>
<td>1267</td>
<td>23.10</td>
<td>3.27</td>
<td>66</td>
<td>22.39</td>
<td>3.89</td>
<td>-1.45</td>
</tr>
<tr>
<td>Gender</td>
<td>1267</td>
<td>0.92</td>
<td>0.27</td>
<td>66</td>
<td>0.88</td>
<td>0.33</td>
<td>-0.97</td>
</tr>
<tr>
<td>Race</td>
<td>1267</td>
<td>0.89</td>
<td>0.31</td>
<td>66</td>
<td>0.85</td>
<td>0.36</td>
<td>-0.97</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01. *** p < .001.
### Table 6

*Characteristic, Demographic, and Test-Score Differences between Male and Female Participants*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Male</th>
<th>Female</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Education</td>
<td>1233</td>
<td>2.61</td>
<td>0.83</td>
</tr>
<tr>
<td>High school grade</td>
<td>1231</td>
<td>2.29</td>
<td>0.63</td>
</tr>
<tr>
<td>Tutor or study group</td>
<td>1219</td>
<td>1.86</td>
<td>1.12</td>
</tr>
<tr>
<td>Practice test</td>
<td>1231</td>
<td>2.74</td>
<td>1.45</td>
</tr>
<tr>
<td>Number times TPM read</td>
<td>1230</td>
<td>2.80</td>
<td>1.41</td>
</tr>
<tr>
<td>Easy to read</td>
<td>1191</td>
<td>3.83</td>
<td>0.85</td>
</tr>
<tr>
<td>Motivated to study</td>
<td>1208</td>
<td>3.88</td>
<td>0.96</td>
</tr>
<tr>
<td>Quiet place to study</td>
<td>1203</td>
<td>3.79</td>
<td>0.99</td>
</tr>
<tr>
<td>Total hours studying</td>
<td>1225</td>
<td>2.72</td>
<td>1.22</td>
</tr>
<tr>
<td>Final two weeks studying</td>
<td>1224</td>
<td>3.10</td>
<td>1.32</td>
</tr>
<tr>
<td>Study self-efficacy</td>
<td>1203</td>
<td>4.03</td>
<td>0.93</td>
</tr>
<tr>
<td>Importance of hire</td>
<td>1233</td>
<td>4.66</td>
<td>0.65</td>
</tr>
<tr>
<td>Firefighter self efficacy</td>
<td>1233</td>
<td>4.87</td>
<td>0.36</td>
</tr>
<tr>
<td>Certain time to study</td>
<td>1213</td>
<td>3.75</td>
<td>1.03</td>
</tr>
<tr>
<td>Hours read on job</td>
<td>1168</td>
<td>4.11</td>
<td>0.96</td>
</tr>
<tr>
<td>TPM test score</td>
<td>1233</td>
<td>52.87</td>
<td>13.85</td>
</tr>
<tr>
<td>Cognitive ability score</td>
<td>1233</td>
<td>23.19</td>
<td>3.25</td>
</tr>
<tr>
<td>Age</td>
<td>1221</td>
<td>27.67</td>
<td>5.71</td>
</tr>
<tr>
<td>Race</td>
<td>1233</td>
<td>0.89</td>
<td>0.31</td>
</tr>
</tbody>
</table>

*p < .05. ** p < .01. *** p < .001.
As can be seen in Table 4, Black participants indicated they had significantly higher high-school grades, were somewhat older, spent more time studying with a tutor or study group while preparing for the TPM test, were more likely to set aside a certain time to study, and were more like to report it was important to be hired as a firefighter by one of the two testing agencies. Blacks also reported spending more total time studying and studied longer during the final two weeks before the tests were administered, than White participants. Black participants scored significantly lower on the TPM and cognitive ability measures than did White participants.

Compared to participants under 40 years of age, participants who were 40 or older reported reading the TPM test significantly more times, being more motivated to study, and studying longer during the total time before the tests were administered (see Table 5). As compared to men, women reported taking the practice test significantly more times, possessing more study self-efficacy, spending more time studying both during the final two weeks and for the total time period before the test was administered, and being more likely to set aside a certain time for study (see Table 6). However, women also reported having significantly lower high school average grades and they scored significantly lower on the cognitive ability measure.

Examination of potential multicollinearity issues

A number of the hypotheses were tested using regression analysis. One assumption of multiple linear regression (MLR) is the lack of extreme multicollinearity in the predictors. This is important because the most efficient prediction in a multiple regression analysis takes place when each predictor is highly correlated to the criterion, but is unrelated to other predictors. Meyers, Gamst, and Guarino (2006) recommended
that “two variables correlated in the middle .7s or higher should probably not be used together in a regression” (page 181). Furthermore, Grimm and Yarnold (1995) indicated “most investigators would probably agree that correlations of $r > .80$ should be considered very problematic” (page 45).

An examination of the correlation matrix (Table 3) determined that only the predictors measuring the total amount of time test takers spent studying the TPM manual before the test was administered and the amount of time studying during the final two weeks before the test was administered were sufficiently correlated to be of potential concern for multicollinearity ($r = .79$). However, because these two variables would have contributed redundant variance in the prediction of TPM test scores and the individual contribution of each of these variables was of specific interest, they were examined separately during the tests of the hypotheses.

There were two primary reasons why the individual contribution of the two study-time intervals was of interest. First, a recall bias has been reported in the literature when self-report data was collected with recent events being recalled more accurately than more distal events (e.g., Catania, Gibson, Chitwood, & Coates, 1990; Kauth, St. Lawrence & Kelly, 1991). Based on that research it was anticipated that participants might be more likely to accurately report the amount of time studied during the final two weeks prior to the test being administered than the amount of time studying for many weeks, or even months, before the test was administered. Second, previous research has suggested that the retention interval between learning and testing can play an important role in test scores, such test scores are likely to be higher when information is learned more recently than for information learned earlier (e.g., Halpern & Hakel, 2003). Thus, it
was anticipated that there might be stronger relationship between more recent studying and test performance than for total time studying and test performance.

*Hypothesis 1*

It was predicted that subgroup mean differences would be lower when a TPM test was used than when using a more traditional measure of cognitive ability. This hypothesis was tested using a repeated measures analysis of variance, which SPSS conducts as an MANOVA. The repeated-measures factors in this analysis were the test takers’ standardized TPM test scores and their standardized scores on the measure of cognitive ability. The between subject factor was race (i.e., Black/White). It should be noted, that in performing this analysis, both sets scores had been standardized or converted to z scores at the level of the total sample. The statistic used to analyze the test performance main effect and race by test performance scores interaction effects was Wilks’ lambda ($\Lambda$). The effect of greatest interest was that for the interaction. The race by test performance interaction effect was significant, $\Lambda = .986, F(1, 1342) = 18.429, p < 001$. An examination of the mean effect-size differences between Black and White test takers found a $d$ of 0.334 on the TPM test and a $d$ of 0.764 on the measure of cognitive ability (see Table 7). In other words, the mean effect-size differences between the test scores were more than twice as great for the cognitive ability measure than for the TPM test, which is in the hypothesized direction. Therefore, Hypothesis 1 was fully supported.

*Hypothesis 2*

This hypothesis predicted that cognitive ability would be positively related to learning as measured by the TPM test. The relationship between cognitive ability and
Table 7

*Mean Effect Size-Score Differences (d) Between Black and White Test Takers on the TPM Test and the Measure of g.*

<table>
<thead>
<tr>
<th></th>
<th>White</th>
<th>Black</th>
<th></th>
<th></th>
<th></th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>N</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>TPM test</td>
<td>1194</td>
<td>53.50</td>
<td>13.536</td>
<td>150</td>
<td>48.88</td>
<td>15.868</td>
</tr>
<tr>
<td>Cognitive ability</td>
<td>1194</td>
<td>23.33</td>
<td>3.092</td>
<td>150</td>
<td>20.87</td>
<td>4.102</td>
</tr>
</tbody>
</table>
TPM test scores was found to be both significant and positive ($r = .379$, $r^2 = .144$, $p < .001$, $n = 1344$). This indicated full support for Hypothesis 2.

It should be noted that cognitive ability was relatively independent of many of the study variables that were of primary interest to the current study. The only significant correlations for cognitive ability to variables that were of primary interest in the current study were for studying with a tutor or study group, the amount of time studied during the final two weeks before the test was administered, study self-efficacy, TPM test score, and race. However, cognitive ability was not significantly related to the number of times a practice test was taken, motivation to study the TPM manual, total hours studying the TPM manual, or setting aside a certain time to study.

*Hypothesis 3a*

This hypothesis predicted that self-reported time studying the TPM manual would be positively related to TPM test scores. The relationship between total hours studying the TPM manual and TPM test scores was found to be both positive and significant ($r = .485$, $r^2 = .235$, $p < .01$, $n = 1336$). Similarly, the relationship between the amount of time studying the TPM manual the final two weeks before the test was administered and TPM test scores was also found to be both positive and significant ($r = .563$, $r^2 = .317$, $p < .01$, $n = 1334$). These findings indicated that Hypothesis 3a was fully supported for both total time studying the TPM manual and for the amount of time the TPM manual was studied the final two weeks before the test was administered.

*Hypothesis 3b*

This hypothesis predicted that time spent studying the TPM manual and cognitive ability would interact to effect TPM test scores such that those with lower levels of
cognitive ability would benefit more from increased study time than would those with higher levels of higher levels of cognitive ability. Hypothesis 3b was tested using MLR to determine whether a significant interaction would exist consistent with the hypothesis that those with lower levels of cognitive ability gained more than those with higher levels of cognitive ability from additional time studied.

A regression was first conducted predicting TPM test scores from cognitive ability, total time spent studying, and the interaction of cognitive ability and total time spent studying. In Step 1, Total time studying and cognitive ability accounted for a significant amount of the variance ($R^2 = .381, p < .001$) (see Table 8). Adding the interaction of total study time and cognitive ability at Step 2 accounted for significant incremental variance ($\Delta R^2 = .008, p < .001$). A graphic representation of the TPM test scores of low and high cognitive ability participants as a function of the total amount of time studying the TPM manual is shown as Figure 4. Although the interaction was significant, Hypothesis 3b was not supported since those with higher levels of cognitive ability benefited more from increased total study time than those with lower levels of $g$, which is in the opposite of the predicted direction. In addition, although the increase in $R^2$ was significant, in practical terms it was very small (i.e., less than .01).

The next regression analysis that was conducted involved the prediction of TPM test scores from cognitive ability, total time spent studying, and the interaction of cognitive ability and total time spent studying. In Step 1, time spent studying during the final two weeks prior to the test being taken and cognitive ability accounted for a significant amount of the variance in the regression analysis ($R^2 = .437, p < .001$; see Table 9). Adding the interaction of the time spent studying during the final two weeks
Table 8

Regressions Analysis Summary for the Interaction of Total Time Studying and Cognitive Ability Predicting TPM Test Score \((N = 1336)\)

<table>
<thead>
<tr>
<th>Variable</th>
<th>(B)</th>
<th>(SEB)</th>
<th>(\beta)</th>
<th>(R^2)</th>
<th>(\Delta R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive ability</td>
<td>.381</td>
<td>.022</td>
<td>.382***</td>
<td>.381***</td>
<td></td>
</tr>
<tr>
<td>Total time studying</td>
<td>.482</td>
<td>.021</td>
<td>.489***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.008***</td>
</tr>
<tr>
<td>Cognitive ability</td>
<td>.382</td>
<td>.021</td>
<td>.382***</td>
<td>.389***</td>
<td>.008***</td>
</tr>
<tr>
<td>Total time studying</td>
<td>.484</td>
<td>.021</td>
<td>.491***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive ability × Total time</td>
<td>.089</td>
<td>.021</td>
<td>.092***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>studying</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Even though the \(\Delta R^2\) was significant in Step 2, the relationship corresponding to the \(\Delta R^2\) was in the opposite direction of that predicted by Hypothesis 3b.

*** \(p < .001\).
Figure 4. TPM test scores of low, mean, and high cognitive ability participants as a function of the total amount of time studying the TPM manual. Computed slopes of regression line for level of cognitive ability and study time interaction are presented. Even though the interaction was significant at the $p < .05$ level, it was in the opposite direction of the prediction made in Hypothesis 3b.
Table 9

*Regression Analysis Summary for the Interaction of Final Two Weeks Studying and Cognitive Ability Predicting TPM Test Score (N = 1334)*

<table>
<thead>
<tr>
<th></th>
<th>Variable</th>
<th>B</th>
<th>SEB</th>
<th>β</th>
<th>R²</th>
<th>ΔR²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cognitive ability</td>
<td>.346</td>
<td>.021</td>
<td>.347***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Final two weeks studying</td>
<td>.539</td>
<td>.020</td>
<td>.543***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td>.352</td>
<td>.021</td>
<td>.353***</td>
<td>.442</td>
<td>.004</td>
</tr>
<tr>
<td></td>
<td>Cognitive ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Final two weeks studying</td>
<td>.537</td>
<td>.020</td>
<td>.542***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cognitive ability × Final two weeks studying</td>
<td>.068</td>
<td>.021</td>
<td>.067**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Even though the ΔR² was significant in Step 2, the relationship corresponding to the ΔR² was in the opposite direction of that predicted by Hypothesis 3b.

**p < .01. ***p < .001.
before testing and cognitive ability at Step 2 accounted for significant incremental variance ($\Delta R^2 = .004, p < .001$). A graphic representation of the TPM test scores of low and high cognitive ability participants as a function of the total amount of time studying the TPM manual is shown as Figure 5. Similar to total time studying, those with higher levels of cognitive ability benefited more from increased study time during the final two weeks than those with lower levels of g. As with total time studying the TPM manual, although the increase in $R^2$ was significant, in practical terms it was very small (i.e., less than .01). Therefore, Hypothesis 3b was also not supported for the time studying the two weeks prior to the test being administered since those with higher levels of cognitive ability benefited more from increased study time than those with lower levels of g, which is the opposite of the predicted direction.

**Hypothesis 4a**

This hypothesis predicted that self-reported levels of motivation to study the TPM manual would be positively related to the amount of time studying the TPM manual. It was determined that motivation to study was significantly related in a positive direction to the total amount of time the TPM manual was studied ($r = .481, r^2 = .231, p < .01, n = 1313$) and to the amount of time the TPM manual was studied during the final two weeks before the test was administered ($r = .531, r^2 = .282, p < .01, n = 1311$), fully supporting Hypothesis 4a.

**Hypothesis 4b**

This hypothesis predicted that self-reported levels of motivation to study the TPM manual would be positively related to TPM test scores. The relationship between these
Figure 5. TPM test scores of low, mean, and high cognitive ability participants as a function of the final two weeks studying the TPM manual. Computed slopes of regression line for level of cognitive ability and study time interaction are presented. Even though the interaction was significant at the p < .05 level, it was in the opposite direction of the prediction made in Hypothesis 3b.
two variables was found to be significantly related in a positive direction ($r = .388$, $r^2 = .150$, $p < .01$, $n = 1318$), fully supporting Hypothesis 4b.

**Hypothesis 4c**

This hypothesis predicted that the amount of time studying the TPM manual would mediate the relationship of self-reported levels of motivation to study the TPM manual and TPM test scores. Mediation answers the question of how an effect occurs. The test of mediation was conducted for both the total amount of reported time studying the TPM manual and the amount of time studying the TPM manual during the final two weeks before the test was administered following the previously described approach.

During the current study the four steps recommended by Baron and Kenny (1986) were followed to test whether mediation occurred as predicted. First, in order for there to be mediation the relationship between independent variable and the mediator must be significant. Second, the relationship between the mediator and dependent variable must be significant. Third, the relationship between independent variable and dependent variable must be significant. Finally, when the mediator and independent variable are used to predict the dependent variable, the significant path between the independent variable and dependent variable must be greatly reduced, if not rendered non significant.

However, even though the fourth step in the process recommended by Baron and Kenny could identify whether the relationship between the independent and dependent variables was reduced when the mediator and independent variable were used to predict the dependent variable, it does not answer the question as to whether the reduction was significant. To address that question in each instance when there was a reduction found in the path between the independent and dependent variables during the fourth step as
described above, a Sobel (1982) test was conducted using the interactive calculation tool offered on the internet by Preacher and Leonardelli (2008).

According to Preacher and Leonardelli (2008) there are three principle versions of the Sobel test. They recommended researchers use the results from the computation of the Aroian (1944/1947) version since it does not make the assumption that the product of the standard error of the raw regression coefficient for the association between the independent and mediator and the standard error of the raw regression coefficient for the association between the mediator and the dependent variable is "increasingly small" (Preacher & Leonardelli, 2008). The following were the steps taken to identify whether significant mediation occurred first for total study time and then for the amount of time studied during the final two weeks before the test was administered.

First, a significant relationship was found between motivation to study (predictor) and total study time (mediator) ($\beta = .481, t = 19.846, r^2 = .231, p < .001$). Second, the relationship between total time studying the TPM manual and TPM test scores (dependent variable) was found to also be significant ($\beta = .485, t = 20.237 r^2 = .235, p < .001$). Third, the relationship between motivation to study and TPM test scores was found to be significant ($\beta = .388, t = 15.274, r^2 = .151, p < .001$). Finally, when motivation to study and total time studying the TPM manual were used to predict TPM test scores, the path between motivation to study and TPM test scores was reduced, however it still remained significant ($\beta = .205, t = 7.542, r^2 = .042, p < .001$).

The Aroian (1944/1947) version of the Sobel (1982) test was conducted which resulted in a finding of $Z = 14.132, p < .001$. This indicated that significant mediation took place. It should be noted that if the association between the independent and
dependent variables had been reduced to a non-significant level then full mediation would have been identified. However, since the association between the predictor and dependent variable was still significant when the mediator was introduced to the regression equation, partial mediation was identified by the Aroian version of the Sobel test. The offered partial support of Hypothesis 4c for the total amount of time test takers reported studying the TPM manual.

Next, a series of regressions were conducted to examine the role of the amount of self-reported time studying the TPM manual during the final two weeks as a mediator on the effects of motivation to study on the prediction to TPM test performance. First, a significant relationship was found between motivation to study (predictor) and time studying the TPM manual during the final two weeks (mediator) ($\beta = .531, t = 22.676, R^2 = .282, p < .001$). Second, the relationship between time studying the TPM manual the final two weeks and TPM test scores (dependent variable) was found to also be significant ($\beta = .563, t = 24.886, R^2 = .317, p < .001$). Third, the relationship between motivation to study and TPM test scores was significant ($\beta = .388, t = 15.274, R^2 = .151, p < .001$). Finally, when motivation to study and the amount of time studying the TPM manual during the final two weeks were used to predict TPM test scores, the path between motivation to study and TPM test scores was smaller, but still significant ($\beta = .129, t = 4.784, R^2 = .017, p < .001$).

The Aroian (1944/1947) version of the Sobel (1982) test was conducted which resulted in a finding of $Z = 16.997, p < .001$. This indicated that significant mediation took place. Since the association between the predictor and dependent variable was still significant when the mediator was introduced to the regression equation, partial
mediation was identified by the Aroian version of the Sobel test. The offered partial support of Hypothesis 4c for the amount of time test takers reported studying the TPM manual the final two weeks before the test was administered.

_Hypothesis 5a_

This hypothesis predicted that self-reported study self-efficacy would be positively related to the amount of time studying the TPM manual. It was determined that study self-efficacy was significantly and positively related to the total amount of time the TPM manual was studied ($r = .350, r^2 = .123, p < .01, n = 1311$) and to the amount of time the TPM manual was studied during the final two weeks before the test was administered ($r = .406, r^2 = .165, p < .01, n = 1308$), fully supporting Hypothesis 6a.

_Hypothesis 5b_

This hypothesis predicted that self-reported levels of study self-efficacy would be positively related to TPM test scores. The relationship between these two variables was found to be positive and significant ($r = .354, r^2 = .125, p < .01, n = 1313$), fully supporting Hypothesis 5b.

_Hypothesis 5c_

This hypothesis predicted that the effect of study self-efficacy on TPM test scores would be mediated by the amount of time the TPM manual was studied. A test of mediation was conducted for both the total amount of reported time studying the TPM manual and the amount of time studying the TPM manual during the final two weeks before the test was administered following the approach previously described in this chapter.
First, a series of regressions were conducted to examine the role of the total time studying the TPM manual as a mediator on the effects of study self-efficacy on the prediction to TPM test performance. A significant relationship was found between study self-efficacy (predictor) and overall study time (mediator) ($\beta = .350$, $t = 13.530$, $R^2 = .123$, $p < .001$). Then, the relationship between total time studying the TPM manual and TPM test scores (dependent variable) was found to also be significant ($\beta = .485$, $t = 20.237$, $R^2 = .235$, $p < .001$). Next, the relationship between study self-efficacy and TPM test scores was significant ($\beta = .354$, $t = 13.689$, $R^2 = .125$, $p < .001$). Finally, when study self-efficacy and total time studying the TPM manual were used to predict TPM test scores, the path between study self-efficacy and TPM test scores was reduced, but still significant ($\beta = .214$, $t = 8.443$, $R^2 = .046$, $p < .001$).

The Aroian (1944/1947) version of the Sobel (1982) test was conducted which resulted in a finding of $Z = 11.226$, $p < .001$. This indicated that significant mediation took place. However since the association between the predictor and dependent variable was still significant when the mediator was introduced to the regression equation, significant partial mediation was identified by the Aroian version of the Sobel test. The offered partial support of Hypothesis 5c for the total amount of time test takers reported studying the TPM manual before the test was administered.

A series of regressions were also conducted to examine the role of time studying the TPM manual during the final two weeks before the test was administered as a mediator on the effects of study self-efficacy on the prediction to TPM test performance. First, a significant relationship was found between study self-efficacy (predictor) and time studying during the final two weeks (mediator) ($\beta = .406$, $t = 16.078$, $r^2 = .165$, $p <$
.001). Second, the relationship between time studying the TPM manual during the final two weeks and TPM test scores (dependent variable) was found to also be significant ($\beta = .563, t = 24.886, r^2 = .317, p < .001$). Third, the relationship between study self-efficacy and TPM test scores was significant ($\beta = .354, t = 13.689, r^2 = .125, p < .001$). Finally, when study self-efficacy and time studying the TPM manual during the final two weeks were used to predict TPM test scores, the path between study self-efficacy and TPM test scores was reduced, but still significant ($\beta = .155, t = 6.202, r^2 = .024, p < .001$).

The Aroian (1944/1947) version of the Sobel (1982) test was conducted which resulted in a finding of $Z = 11.226, p < .001$. This indicated that significant mediation took place. However since the association between the predictor and dependent variable was still significant when the mediator was introduced to the regression equation, significant partial mediation was identified by the Aroian version of the Sobel test. The offered partial support of Hypothesis 5c for the amount of time test takers reported studying the TPM manual the final two weeks before the test was administered.

**Hypothesis 6**

This hypothesis predicted that the effect of race on TPM test scores would be mediated by cognitive ability. To test hypothesis 6 a series of regressions were conducted to examine the role of cognitive ability as a mediator on the effects of race on the prediction to TPM test performance.

First, a significant relationship was found between race (predictor) and cognitive ability (mediator) ($\beta = .234, r^2 = .055, t = 8.834, p < .001$). Second, the relationship between cognitive ability and TPM test scores (dependent variable) was found to also be significant ($\beta = .379, r^2 = .143, t = 14.989, p < .001$). Third, the relationship between race
and TPM test scores was significant ($\beta = .105, r^2 = .011, t = 3.860, p < .001$). Finally, when race and cognitive ability were used to predict TPM test scores, the significant path between race and TPM test scores was no longer significant ($\beta = .017, r^2 = .000, t = .652, p = .515$).

The Aroian (1944/1947) version of the Sobel (1982) test was conducted which resulted in a finding of $Z = 7.637, p < .001$. This indicated that significant mediation took place. Since the association between the predictor and dependent variable was not significant when the mediator was introduced to the regression equation, significant full mediation was identified by the Aroian version of the Sobel test. This offered full support of Hypothesis 6.

**Research Question 1**

This asked whether time spent studying the TPM manual and race would interact to affect TPM test scores such that Black test takers benefited more from increased study time than White test takers. This question was tested using MLR to determine whether significant interactions between race and study time in the direction predicted would be detected. The dependent variable used during this analysis was the test takers’ standardized TPM test scores.

Race accounted for a significant amount of the variance in Step 1 of the regression analysis ($R^2 = .012, p < .001$) (see Table 10). Adding race and total study time at Step 2 accounted for significant additional variance ($\Delta R^2 = .249, p < .001$). Finally, the interaction of total study time and race at Step 3 did not account for significant incremental variance ($\Delta R^2 = .002, p = .099$). Thus, in answer to Research Question 1,
Table 10

*Regression Analysis Summary for the Interaction of Total Time Studying and Race Predicting TPM Test Score (N = 1336)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SEB</th>
<th>β</th>
<th>(R^2)</th>
<th>(\Delta R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td>.012***</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>.344</td>
<td>.085</td>
<td>.110***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td>.261***</td>
<td>.249***</td>
</tr>
<tr>
<td>Race</td>
<td>.506</td>
<td>.074</td>
<td>.161***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total time studying</td>
<td>.494</td>
<td>.023</td>
<td>.501***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 3</td>
<td></td>
<td></td>
<td></td>
<td>.262</td>
<td>.002</td>
</tr>
<tr>
<td>Race</td>
<td>.535</td>
<td>.076</td>
<td>.171***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total time studying</td>
<td>.592</td>
<td>.064</td>
<td>.601***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race × total time studying</td>
<td>-.113</td>
<td>.069</td>
<td>-.106</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The race by total study-time interaction was non significant \( (p = .099)\).

*** \(p < .001\).
Black test takers did not benefit significantly more than White test takers from additional total time studying the TPM manual.

A similar analysis was conducted for time spent studying during the final two weeks prior to the test being taken and race. Once again, race accounted for a significant amount of the variance in Step 1 of the regression analysis ($R^2 = .011, p < .001$) (see Table 11). Adding race and study time during the final two weeks at Step 2 accounted for significant additional variance ($\Delta R^2 = .326, p < .001$). Finally, the interaction of the time spent studying during the final two weeks and race at Step 3 accounted for significant incremental variance ($\Delta R^2 = .002, p = .048$) in the hypothesized direction. The predicted values for level of race and study-time interaction were computed using the obtained regression weights and a graphic presentation of this is presented in Figure 6. In answer to Research Question 1 Black test takers did benefit slightly, but significantly more, than White test takers from additional total time studying the TPM manual during the final two weeks before the test was administered. It should be noted, however, that even though the increase in $R^2$ was significant, in practical terms it was very small (i.e., less than .01).

**Research Question 2**

This asked whether Black-White subgroup mean differences and potential adverse impact on the TPM test would decrease as the amount of time studying the TPM test increased. In order to help better understand the relationships found during the examination of Research Question 1, the means and standard deviations for the TPM test scores at each of the five study-time intervals for the total time studying and the time studying during the final two weeks are reported in Tables 12 and 13 respectively.
Table 11

Regression Analysis Summary for the Interaction of Final Two Weeks Studying and Race Predicting TPM Test Score (N = 1334)

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SEB</th>
<th>β</th>
<th>R²</th>
<th>ΔR²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>0.327</td>
<td>0.086</td>
<td>0.104***</td>
<td>.011***</td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>0.442</td>
<td>0.070</td>
<td>0.141***</td>
<td>.337***</td>
<td>.326***</td>
</tr>
<tr>
<td>Final two weeks studying</td>
<td>0.567</td>
<td>0.022</td>
<td>0.572***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>0.465</td>
<td>0.071</td>
<td>0.148***</td>
<td>.339*</td>
<td>.002*</td>
</tr>
<tr>
<td>Final two weeks studying</td>
<td>0.691</td>
<td>0.066</td>
<td>0.697***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race × Final two weeks studying</td>
<td>-0.139</td>
<td>0.070</td>
<td>-0.132*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05. *** p < .001.
Figure 6. Black and White participants’ TPM test scores as a function of the amount of time studying the TPM manual during the final two weeks before the test was administered. Computed slopes of regression line for level of race and study time interaction are presented.
Table 12

Mean Effect-Size Differences between Black and White TPM Test Scores at Various Total Study-Time Intervals

<table>
<thead>
<tr>
<th>Total study time (hours)</th>
<th>White</th>
<th></th>
<th></th>
<th>Black</th>
<th></th>
<th></th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>162</td>
<td>39.34</td>
<td>8.944</td>
<td>18</td>
<td>29.83</td>
<td>5.113</td>
<td>1.099</td>
</tr>
<tr>
<td>1-15</td>
<td>461</td>
<td>51.24</td>
<td>12.013</td>
<td>39</td>
<td>43.82</td>
<td>14.973</td>
<td>0.605</td>
</tr>
<tr>
<td>16 - 30</td>
<td>269</td>
<td>57.39</td>
<td>11.275</td>
<td>33</td>
<td>49.55</td>
<td>10.244</td>
<td>0.702</td>
</tr>
<tr>
<td>31 - 45</td>
<td>154</td>
<td>59.38</td>
<td>12.429</td>
<td>27</td>
<td>51.93</td>
<td>13.909</td>
<td>0.589</td>
</tr>
<tr>
<td>46 +</td>
<td>141</td>
<td>63.82</td>
<td>11.968</td>
<td>32</td>
<td>61.78</td>
<td>14.491</td>
<td>0.164</td>
</tr>
</tbody>
</table>
Table 13

Mean Effect-Size Differences Between Black and White TPM Test Scores at Various Study-Time Intervals During the Final Two Weeks Before the Test was Administered

<table>
<thead>
<tr>
<th>Study time during final two weeks (hours)</th>
<th>White</th>
<th>Black</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>0</td>
<td>167</td>
<td>38.81</td>
</tr>
<tr>
<td>1-5</td>
<td>256</td>
<td>48.52</td>
</tr>
<tr>
<td>6-10</td>
<td>272</td>
<td>53.77</td>
</tr>
<tr>
<td>11-20</td>
<td>282</td>
<td>59.02</td>
</tr>
<tr>
<td>21 +</td>
<td>209</td>
<td>63.59</td>
</tr>
</tbody>
</table>
The mean Black-White TPM test effect-size differences \((d)\) at each level of reported total study amount are also reported in Table 12 and for each study amount during the final two weeks before the test was administered in Table 13 as a preliminary step to answering Research Question 2. The \(d\) statistic is the majority group (i.e., White) mean minus the minority-group (i.e., Black) mean divided by the sample-weighted average of the group standard deviations. Plots of the mean effect-size differences on the TPM test for both total study time and study time during the final two weeks before the test was administered are shown in Figures 7 and 8 respectively.

For comparison purposes, the mean Black-White effect-size differences on the measure of cognitive ability at each level of reported total study amount are reported in Table 14 and for each study amount during the final two weeks before the test was administered are reported in Table 15. Also for comparison purposes plots showing the mean-effect size differences on the measure of cognitive ability for both total study time and study time during the final two weeks before the test was administered are presented in Figures 9 and 10 respectively.

It should be noted that the effect-size differences would be identical no matter whether standardized or raw scores were used during the calculation process. With that in mind, raw scores are shown in Tables 12, 13, 14, and 15 for easier interpretation. However, mean effect-size differences in test scores between groups do not necessarily reflect actual adverse impact. Section 4D of the federal Guidelines (1978) state “a selection rate for any race, sex, or ethnic group which is less than four-fifths \((4/5)\) (or eighty percent) of the rate for the group with the highest rate will generally be
Figure 7. Mean effect-size differences ($d$) in TPM test scores as a function of the total number of hours studying the TPM manual before the test was administered.
Figure 8. Mean effect-size differences ($d$) in TPM test scores as a function of the hours studying the TPM manual the final two weeks before the test was administered.
Table 14

*Mean Effect-Size Differences Between Black and White Scores on the Measure of Cognitive Ability at Various Total Study-Time Intervals*

<table>
<thead>
<tr>
<th>Total study time (hours)</th>
<th>White</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Black</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>d</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>162</td>
<td>23.12</td>
<td>3.464</td>
<td>18</td>
<td>21.17</td>
<td>4.274</td>
<td>0.549</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-15</td>
<td>461</td>
<td>23.37</td>
<td>2.991</td>
<td>39</td>
<td>20.28</td>
<td>3.927</td>
<td>1.006</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 - 30</td>
<td>269</td>
<td>23.50</td>
<td>3.060</td>
<td>33</td>
<td>20.58</td>
<td>4.657</td>
<td>0.894</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31 - 45</td>
<td>154</td>
<td>23.26</td>
<td>2.856</td>
<td>27</td>
<td>20.85</td>
<td>4.130</td>
<td>0.784</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46 +</td>
<td>141</td>
<td>23.25</td>
<td>3.256</td>
<td>32</td>
<td>21.59</td>
<td>3.688</td>
<td>0.497</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total study time (hours)</td>
<td>White</td>
<td></td>
<td>Black</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>d</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>167</td>
<td>22.99</td>
<td>3.485</td>
<td>16</td>
<td>21.00</td>
<td>4.733</td>
<td>0.552</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-5</td>
<td>256</td>
<td>23.21</td>
<td>3.047</td>
<td>26</td>
<td>20.00</td>
<td>4.128</td>
<td>1.016</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-10</td>
<td>272</td>
<td>23.39</td>
<td>2.905</td>
<td>31</td>
<td>19.26</td>
<td>4.740</td>
<td>1.317</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-20</td>
<td>282</td>
<td>23.45</td>
<td>3.221</td>
<td>39</td>
<td>21.03</td>
<td>3.794</td>
<td>0.735</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21+</td>
<td>209</td>
<td>23.57</td>
<td>2.845</td>
<td>36</td>
<td>22.47</td>
<td>2.971</td>
<td>0.384</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 9. Mean effect-size differences ($d$) in scores on the measure of cognitive ability as a function of the total number of hours studying the TPM manual before the test was administered.
Figure 10. Mean effect-size differences ($d$) in scores on the measure of cognitive ability as a function of the hours studying the TPM manual the final two weeks before the test was administered.
regarded by the Federal enforcement agencies as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded by Federal enforcement agencies as evidence of adverse impact. Smaller differences in selection rate may nevertheless constitute adverse impact, where they are significant in both statistical and practical terms” (page 24-103). With this in mind, tests of both the 80% rule of thumb and of statistical differences in the selection rates of Blacks and White test takers were conducted when answering Research Question 2. However, only statistically significant differences will be discussed since the 80% rule is not typically recognized by the courts as the appropriate method for determining adverse impact.

In order to examine differences in selection rates it must first be determined who “passed” (i.e., selected) or “failed” (i.e., rejected) the tests being analyzed. During the validation study of the TPM test SMEs from the two fire departments described in the current study that used the tests for selection purposes had indicated their estimates of the percentage of minimally qualified applicants they believed would answer each test item correctly from both the TPM test and from the measure of cognitive ability (S. Bell, personal communication, April 18, 2007). In other words, each test item was assigned a percentage value by the SMEs. These ratings were averaged and cutoff scores were determined. This is traditionally referred to as the Angoff method for setting cutoff scores (Angoff, 1971). The unmodified average of the suggested cutoff scores were then used to determine selection (i.e., passing) rates for the TPM test (see Tables 16 and 17) at all study-amount levels.

The adverse impact analyses for the passing rates at each of the time intervals were computed using the Adverse Impact Toolkit program (AI Toolkit; Biddle, 2006).
The exact statistical significance test conducted by the AI Toolkit program used a two-tail Fisher Exact probability statistic for 2 X 2 contingency tables that provided standard deviation values of the differences in passing rates between Black and White test takers. A standard deviation value of 1.96 from this test corresponds with a (two-tail) probability value of .05. Values less than 1.96 are statistically significant and indicate a difference in passing rates between two groups that did not likely occur by chance.

As can be seen in Table 16, instead of statistically-based adverse impact (i.e. differences in passing rates) against Blacks on the TPM test systematically decreasing as reported total time studying increased, the relationship was non-linear. For example, there was no statistically significant adverse impact against Blacks on the TPM test found for those test takers who reported not studying at all. However, the failure to find significant adverse impact at this study-time interval appears to be a statistical artifact due to the small number of those who passed the test (ten out of 173 White test takers and none of the 19 Black test takers).

In addition, there was also no statistically significant adverse impact on the TPM test against Blacks for those who reported studying between 1 and 15 hours and who reported studying more than 46 hours in total. Unexpectedly, the strongest significant adverse impact against Blacks occurred for those who reported studying between 16 and 30 hours overall. Also somewhat smaller level of significant adverse impact was found for those who reported studying a total of between 31 and 45 hours.

The relationship between the amount of time studied during the final two weeks before the test was administered and levels of adverse impact against Blacks on the TPM
Table 16

Adverse Impact on TPM Test Between Black and White Test Takers at Varying Amounts of Total Study Time

<table>
<thead>
<tr>
<th>Total study time (hours)</th>
<th>Sample Size</th>
<th>TPM Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Black</td>
<td>White</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>1 - 15</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>16 - 30</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>31 - 45</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>46 +</td>
<td>33</td>
</tr>
</tbody>
</table>

Notes: The 80% Test is an analysis that compares the passing rate of one group to the passing rate of another group (e.g., men vs. women). An 80% Test “violation” occurs if one group’s passing rate is less than 80% of the group with the highest rate.

Standard Deviations describe the probability value (from the statistical test) in terms of standard deviation units. A standard deviation of 1.96 corresponds with a (two-tail) probability value of .05.

*p < .05. **p < .01.
test was also not linear (see Table 17). Again, there no significant adverse impact against Black test takers for those who did not study at all, which again appears to be an artifact of the small number of those who actually passed the test. In addition, there was no significant adverse impact against Blacks for those who reported studying 1 to 5 hours. However, while there was significant adverse impact against those who reported studying 6 to 10 and 11 to 20 hours the two weeks before the test was administered, there was no statistically significant adverse impact against Blacks for those who reported studying the most (i.e., more than 20 hours) during the final two weeks before the test was administered; with the passing rate in this group being higher for Blacks than for Whites.

Thus, the testing of Research Question 2 showed that the least adverse impact took place for those Black and White test takers who studied the most before the test was administered. However, these findings should be interpreted with caution since the number of Black participants was relatively small in each of the study-time increments. So, in summary, the lowest level of adverse impact against Black test takers on the TPM test occurred for those who reported studying the longest time overall and for those who reported studying the most during the final two weeks prior to the test being administered. Even more encouraging in regards to the main thrust of the current study was that the passing rate for Blacks was actually slightly higher than the passing rates for Whites for those who reported studying the most during the final two weeks before the test was administered.

Finally, the level of adverse impact on the TPM test and the measure of cognitive ability for all of the Black and White test takers were calculated. When using the
Table 17

*Adverse Impact on TPM Test Between Black and White Test Takers at Varying Amounts of Study Time During the Final Two Weeks Before the Tests Were Administered*

<table>
<thead>
<tr>
<th>Total study time (hours)</th>
<th>Sample Size</th>
<th>TPM Test</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Black</td>
<td>White</td>
<td>Black Passing Rate</td>
<td>White Passing Rate</td>
<td>80% Test</td>
<td>Standard Deviations</td>
</tr>
<tr>
<td>0</td>
<td>17</td>
<td>178</td>
<td>0%</td>
<td>05.6%</td>
<td>0.00</td>
<td>0.517</td>
</tr>
<tr>
<td>1 - 5</td>
<td>28</td>
<td>265</td>
<td>21.4%</td>
<td>28.3%</td>
<td>0.76</td>
<td>0.656</td>
</tr>
<tr>
<td>6 - 10</td>
<td>40</td>
<td>284</td>
<td>22.5%</td>
<td>46.8%</td>
<td>0.48</td>
<td>2.905**</td>
</tr>
<tr>
<td>11 - 20</td>
<td>40</td>
<td>289</td>
<td>25.0%</td>
<td>63.0%</td>
<td>0.40</td>
<td>4.427***</td>
</tr>
<tr>
<td>21 +</td>
<td>38</td>
<td>217</td>
<td>81.6%</td>
<td>77.9%</td>
<td>1.05</td>
<td>NR^a</td>
</tr>
</tbody>
</table>

*Notes:* The 80% Test is an analysis that compares the passing rate of one group to the passing rate of another group (e.g., men vs. women). An 80% Test “violation” occurs if one group’s passing rate is less than 80% of the group with the highest rate.

Standard Deviations describe the probability value (from the statistical test) in terms of standard deviation units. A standard deviation of 1.96 corresponds with a (two-tail) probability value of .05.

**p < .01. ***p < .001.

^aThe Adverse Impact Toolkit program does not report standard deviations when the selection rate for Blacks is favored over Whites.
unmodified Angoff cutoff score for determining who passed or failed the TPM test, about 36% \((n = 56)\) of Black test takers and 46% \((n = 570)\) of the White test takers passed, which resulted in an adverse impact probability value of 2.394 standard deviations. In addition, about 61% \((n = 96)\) of the Black test takers and 87% \((n = 1065)\) of the White test takers passed the measure of cognitive ability, which resulted in an adverse impact probability value of 7.195 standard deviations.

In order to determine whether the lower adverse impact findings were an artifact because of the difference in the proportion of Black and White test takers, additional analyses were conducted on those Black and White test takers who scored the highest on the TPM test \((n = 303)\) and the cognitive ability test \((n = 302)\). Approximately 23% \((n = 274)\) of the White test takers and 19% \((n = 29)\) of the Black test takers scored 66 or above on the TPM test, which resulted in an adverse impact probability value of 2.027 standard deviations. Approximately 24% \((n = 291)\) of the White test takers and 7% \((n = 11)\) of the Black test takers scored 26 or above on the measure of cognitive ability, which resulted in an adverse impact probability value of 5.449. These findings show that the differences in the pass and fail rates by race (i.e., adverse impact) were much lower for the TPM test than for the measure of cognitive ability.

**Research Question 3**

This research question asked whether a combination of study techniques (including setting aside certain times to study, studying with a tutor or study group, and the number of times a practice test was taken) would explain variance beyond that explained by study time and cognitive ability. It was tested using MLR first for total time
the TPM manual was studied, followed by the amount of time the TPM manual was studied during the final two weeks before the test was administered.

Total time studying the TPM manual and cognitive ability accounted for a significant amount of the variance to the prediction of TPM test scores in Step 1 of the regression analysis ($R^2 = .376, p < .001$) (see Table 18). Adding setting aside certain times to study, studying with a tutor or study group, and the number of times a practice test was taken accounted for significant incremental variance ($\Delta R^2 = .023, p < .001$) in Step 2. This indicated that study strategies predicted TPM test scores above and beyond the total time the TPM manual was studied and cognitive ability. However, even though the increase in the variance of the prediction was significant, it was very small (approximately two percent).

Time studying during the final two weeks before the test was administered and cognitive ability accounted for a significant amount of the variance to the prediction of TPM test scores in Step 1 of the regression analysis ($R^2 = .439, p < .001$) (see Table 19). Adding setting aside certain times to study, studying with a tutor or study group, and the number of times a practice test was taken accounted for significant incremental variance ($\Delta R^2 = .010, p < .001$) in Step 2. This indicated that study strategies predicted TPM test scores above and beyond the total time the amount of time the TPM manual was studied during the final two weeks before the test was administered and cognitive ability. However, even though the increase in the variance of the prediction was significant, it was very small (approximately one percent).
Table 18

*Regression Analysis Summary for Study Strategy Characteristics Predicting TPM Test Scores When Holding Total Time Studying and Cognitive Ability Constant (N = 1290)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SEB</th>
<th>β</th>
<th>R^2</th>
<th>ΔR^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive ability</td>
<td>0.388</td>
<td>0.022</td>
<td>0.394***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total time studying</td>
<td>0.466</td>
<td>0.022</td>
<td>0.470***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td>0.011</td>
<td>0.021</td>
<td></td>
<td>0.400***</td>
<td>0.023***</td>
</tr>
<tr>
<td>Cognitive ability</td>
<td>0.382</td>
<td>0.022</td>
<td>0.389***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total time studying</td>
<td>0.360</td>
<td>0.029</td>
<td>0.364***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certain time to study</td>
<td>0.153</td>
<td>0.026</td>
<td>0.148***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tutor or study group</td>
<td>-0.013</td>
<td>0.024</td>
<td>-0.013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practice test</td>
<td>0.074</td>
<td>0.025</td>
<td>0.075**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05. **p < .01. *** p < .001.
Table 19

Regression Analysis Summary for Study Strategy Characteristics Predicting TPM Test Scores When Holding Time Studying During the Final Two Weeks and Cognitive Ability Constant (N = 1302)

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>B</th>
<th>SEB</th>
<th>β</th>
<th>( R^2 )</th>
<th>( \triangle R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cognitive ability</td>
<td>0.353</td>
<td>0.021</td>
<td>0.357***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time studied final two weeks</td>
<td>0.534</td>
<td>0.021</td>
<td>0.537***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Cognitive ability</td>
<td>0.357</td>
<td>0.021</td>
<td>0.360***</td>
<td>0.439***</td>
<td>.010***</td>
</tr>
<tr>
<td></td>
<td>Time studied final two weeks</td>
<td>0.455</td>
<td>0.027</td>
<td>0.457***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Certain time to study</td>
<td>0.099</td>
<td>0.025</td>
<td>0.096***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tutor or study group</td>
<td>0.008</td>
<td>0.023</td>
<td>0.008</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Practice test</td>
<td>0.050</td>
<td>0.024</td>
<td>0.050*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* \( p < .05 \). ** \( p < .01 \). *** \( p < .01 \).
CHAPTER V
DISCUSSION

One of the goals of the current study was to demonstrate that an entry-level firefighter TPM test would result in significantly lower subgroup mean differences than a traditional measurement of cognitive ability for the selection of entry-level firefighters. The primary goal was to develop a model of TPM performance, as a vehicle for better understanding the variables that might have contributed to the reduced subgroup differences. The following is a discussion of how these two goals were addressed during the current study.

Comparison of Subgroup Differences by Race on the TPM and the Cognitive Ability Measure

As predicted in Hypothesis 1, the Black-White standardized mean-difference on the TPM test \(d = .334\) was significantly lower than on the measure of cognitive ability \(d = .764\), when both tests were administered during the same session to job applicants for a firefighter position. In other words, the mean difference on the TPM test was less than one-half of the mean difference on a traditional measure of cognitive ability. The Black-White mean difference on the TPM test was also much lower than the generally-accepted one standard deviation Black-White difference that has been reported when cognitively-loaded measures are used (Hunter & Hunter, 1984; Roth, BeVier, Bobko, Switzer, & Tyler, 2001).
Furthermore, the adverse impact probability value of 2.394 standard deviations against Black test takers on the TPM test was less than one third of the adverse impact probability value of 7.195 standard deviations on the measure of cognitive ability. Both the lower subgroup mean difference and the lower adverse impact when using the TPM test than when using the measure of cognitive ability are very promising results, in that they suggest that the TPM test provides an alternative to traditional cognitive ability testing.

*Role of Cognitive Ability and Study Time in Predicting TPM Test Scores*

Given that the TPM test resulted in smaller subgroup differences than the cognitive ability test, the next question was why? One possibility was that the TPM test suffered from poor reliability. However, previous research suggested that TPM testing results in high levels of internal consistency and in the current study the obtained reliability based upon coefficient alpha was 0.936 ($n = 2,195$).

The second possibility was the “natural” interaction hypothesis (Asher & Sciarrino, 1974; Smith 1991). That is, according to the natural interaction hypothesis, a test such as the TPM would likely have less adverse impact than a cognitive ability test because it allowed for a natural interaction of cognitive ability and the amount of time the TPM manual was studied. If we assume that the amount of time studying is unrelated to race (or that Black test takers study more than White test takers), and that cognitive ability and study time contribute roughly equally to TPM performance, then it follows that the TPM test should have lower subgroup differences than a traditional measure of cognitive ability.
This natural interaction hypothesis can be modeled simply as presented in Figure 1. In Figure 11 we see that TPM performance is a function of a learning strategy variable (i.e., the amount of time spent studying) and cognitive ability.

The simple model presented in Figure 11 was supported by the analyses. (Note that for simplicity’s sake I am reporting only the amount of time studied during the final two weeks here, but the results were similar for total time studying). Cognitive ability was significantly correlated \((r = .38, p < .001)\) (see Table 3) with TPM scores, which supports Hypothesis 2. The amount of time studied during the final two weeks was also significantly correlated with TPM test performance \((r = .56, p < .001)\). This supports Hypothesis 3b, along with indicating that the amount of time that the TPM manual was studied was more important to the prediction of TPM performance than cognitive ability.

Furthermore, the correlation between cognitive ability and time studied during the final two weeks was significant \((r = .06, p < .05)\), but the amount of variance explained was small, less than one percent. In addition, it should be noted that Black test takers \((M = .336, SD = 1.31, n = 148)\) reported significantly longer mean study times than did White test takers \((M = .309, SD = 1.31, n = 1186; t = .232, p < .05)\). In practical terms, this difference in study times was small, but it further supports the view that there is less adverse impact for the TPM test than for the cognitive ability test because of the contribution of study time.

Thus, support is found for the natural interaction hypothesis and for the Simple Model. Both study time and cognitive ability predicted TPM performance. From a practical perspective, the correlation between TPM performance and cognitive ability
Figure 11. Simple two-factor model for predicting TPM test performance.

Note: For simplicity’s sake, only the Beta weight for time the TPM manual was studied during the final two weeks is shown in this figure.
was so low, they could be considered to be independent. Together, cognitive ability and study time explained 44% of the variance in TPM scores, equivalent to a correlation of $r = .661$ ($p < .001$). If we correct for attenuation, the estimated correlation would be $r = .683$.

According to Hypothesis 3b, the Simple Model involved not only a natural interaction but also a statistical interaction between cognitive ability and study time in predicting TPM scores. That is, it was hypothesized that the effect of self reported study time on TPM test scores would be moderated by cognitive ability, such that those with lower levels of cognitive ability were anticipated to benefit more from additional time studying than those with higher levels of cognitive ability. This hypothesis was based upon Walczyk’s (1993; 2000) Compensatory-Encoding Model, which proposed that low ability learners could compensate through the use of additional study techniques, such as studying longer. Furthermore, Walczyk and Griffith-Ross (2007) indicated that high ability learners were likely to become bored and lose focused as they studied longer, whereas low ability learners were likely to become more engaged. This, they proposed, would result in high-ability learners using fewer strategies than low ability learners as the amount of time studying increased for both groups.

However, Hypothesis 3b was not supported. Even though the interaction of cognitive ability and study time was statistically significant for both total time studying and time studying during the final two weeks, it was in the opposite direction of what was predicted, with high ability test takers benefiting more from additional time studying than low ability test takers. Furthermore, the interactions resulted in an increase of less than
one percent of additional variance in the prediction of TPM scores for both of the study-time variables. Such a small increase was judged not to be practically significant.

**Effect of Race on TPM Scores as Mediated by Cognitive Ability**

Race was correlated with both the total time studying the TPM manual ($r = -.103$, $p < .001$) and the time studied during the final two weeks before the test was administered ($r = -.064$, $p < .05$) (see Table 3). However, the variance accounted for in each of these relationships was about one percent or less, which is so low as not to be practically significant.

Furthermore, as predicted by Hypothesis 6, cognitive ability fully mediated the relationship between race and TPM test performance. In other words, cognitive ability explained the differences in TPM test scores by race. This was in line with the preponderance of the literature (e.g., Jensen, 1985; Gottfredson, 2002, Rushton, 2003). which has reported that cognitive ability generally serves as the primary explanation for Black-White mean differences when cognitively-loaded measures are used.

**Interaction of Race and Study Time**

A test of Research Question 1 revealed that Black test takers benefited more from additional time studying the TPM manual during the final two weeks before the test was administered than White test takers. At first glance this suggested that something was unique about TPM testing that benefited Black test takers over White test takers. However, the importance of this finding was mitigated by three factors. First, the current study found that Black test takers reported studying significantly longer than White test takers for both the total time before the test was administered and for the two weeks before the test was administered. This would appear to be in line with the research of
those such as Olivares (2002) and Durand and Rau (2000), who indicted that low ability learners typically study longer than high ability learners. Since time studying the TPM manual was significantly related to TPM test performance, the lower Black-White mean score difference in TPM test performance may have been at least partially due to Black test takers studying longer than White test takers.

Next, the test of Research Question 1 revealed the interaction of race and time studying during the final two weeks increased the variance of the prediction of TPM test performance over study time during the final two weeks and race by only about two-tenths of one percent ($\Delta R^2 = 0.002, p < .05, n = 1334$), which is extremely small given the number of participants that were tested. As a result, this interaction was viewed as not reaching a level indicating practical significance. Also, the race and time studying during the final two weeks interaction was no longer significant when a post hoc analysis was conducted with cognitive ability being held constant. Third, the effect was not significant for the total amount of time the TPM manual was studied. Thus, Black test takers did not appear to practically benefit more from additional time studying the TPM manual than did White test takers.

However, as predicted, the Black-White mean effect-size differences in TPM test scores did generally decrease as the total amount of study time increased (see Figure 7). In other words, as both Black and White test takers studied more the standardized-mean differences between their scores diminished – that is, their scores on the TPM test became more similar. This is extremely important since previous research had suggested that test preparation was not very likely to help minimize adverse impact on cognitively-
Finally, it should be noted that the pattern of the Black-White mean effect size differences in TPM test scores based on the amount of time studied during the final two weeks before the test was administered was much less linear than originally anticipated (see Figure 8). Even so, the largest mean effect-size differences for both of the study-time interval conditions (i.e., total study time and the amount of time studied during the final two weeks before the test was administered) occurred for those who did not study the TPM manual at all and the lowest mean effect-size differences occurred for those who studied the most for both total time studying and time studied during the final two weeks. This suggested that study time did play an important role in the reduction of the effect-size differences between Black and White test takers, but that this effect was not as linear as expected. More research into this phenomenon is strongly encouraged.

*Effects of Motivation and Self-Efficacy*

According to the more complex model of TPM performance (see Figure 2), the effect of race would be mediated by cognitive ability and study time would be explained by motivation to study and self efficacy. Both motivation to study and study self-efficacy were measured using one item questions. Aspects of the relationships predicted in the more complex TPM performance model are graphically reported in Figure 12.

As reported above, a test of Hypothesis 6 revealed that cognitive ability fully mediated the relationship between TPM performance and race. Furthermore, as seen in Figure 12, when motivation to study and study self-efficacy were entered during the same
Figure 12. Selected aspects of a model of the relationships (beta) between components related to performance on an entry-level firefighter TPM test.

Note: For simplicity’s sake only relationships related to time studying during the final two weeks are presented here. The relationships for the total amount of time studying the TPM manual are similar. Because of the small amount of incremental variance explained by motivation to study and study self-efficacy on the prediction of TPM performance, the figure is shown as if the prediction of TPM performance by motivation to study and study self-efficacy were fully mediated by the amount of time studying to make the final model more parsimonious.
step for the prediction of TPM performance, motivation to study significantly predicted the amount of time studying during the final two weeks before the test was administered ($\beta = .425, p < .001$), which supports Hypothesis 4a. Similarly, when motivation to study and study self-efficacy were entered during the same step in the prediction of TPM performance study self-efficacy significantly predicted the amount of time studying during the final two weeks before the test was administered ($\beta = .197, p < .001$), which supports Hypothesis 5a. Furthermore, the relationship between motivation to study and TPM performance, and the relationship between study self-efficacy and TPM performance, were both partially mediated by the amount of time the TPM manual was studied, which offered partial support to Hypotheses 4c and 5c. Thus, motivation to study and study self-efficacy were found to indirectly aid in the prediction of TPM performance through the amount of time the TPM was studied.

Interestingly, Black and White test takers reported possessing similar levels of motivation to study in the current study (see Table 4). This finding ran counter to those of Ryan, Ployhart, Greguras, & Schmit (1998) who reported that Black firefighter job applicants in their study indicated lower motivation levels than White applicants ($t = -2.16, p = .03$). Other studies have also reported White test takers being more motivated than Blacks to perform well on tests (e.g., Arvey, Strickland, Draude, & Martin, 1990; Chan, Schmitt, DeShon, Clause, & Delbridge, 1997).

Along similar lines to motivation, there was also no significant difference in the study self-efficacy of Black and White test takers in the current study (see Table 4). This again was different from the study by Ryan et al. (1998), who indicated Black job applicants reported statistically significant lower levels of what they called firefighter
self-efficacy (i.e., smart enough to be a firefighter) than White applicants ($F(1,955) = 11.54, p < .001$). Even though Ryan et al. measured firefighter self-efficacy and not study self-efficacy, it appeared that study self-efficacy in the current study could serve as a proxy for firefighter self-efficacy in Ryan et al.’s study, since the job applicants in the current study were told that reading and studying the TPM manual was very similar to important tasks performed by firefighters both in the academy and on the job.

So, why were differences in motivational beliefs by Black and White job applicants different in the current study when compared to other studies? One possible explanation was that a greater number of test takers with low levels of motivational beliefs might have withdrawn from the selection process prior to taking the TPM test than might occur when more traditional types of testing are offered. TPM testing is unique in that it allows test takers to read and study printed material for weeks in advance of the test, whereas more traditional testing methods provide relatively little time for an applicant to reflect about how motivated they might be in regards to the test. The practice test that was based on the TPM manual would also help to encourage those who were motivated and who performed well on that practice test. Thus, it would appear that the TPM testing process would provide a greater impetus for those with lower levels of motivation to withdraw, thereby resulting in a more homogenous group being tested.

Also, job applicants in the current study were informed well in advance of the test that a minimum requirement to be offered a position was the possession of a basic EMT certification. It is likely that motivation to study would have played a role in success during that certification process, just as it would have in the current study. Thus, the EMT certification process could have served as a tool to screen out unmotivated potential job
applicants, resulting in the remaining test takers being more similar in their level of motivation than if they had not been pre-screened by this requirement.

The findings about the relationships between the individual-difference variables and TPM test performance in the current study were similar to those that have been generally reported in the literature. More specifically, motivation to study the TPM manual and study self-efficacy were both positively related to learning as measured by the TPM test. Furthermore, motivation to study and study self-efficacy were positively related to the amount of time the TPM manual was studied, which was then itself related to TPM test performance. Next, the relationship between motivation to study and TPM performance, and the relationship between study self-efficacy and TPM performance, were partially mediated by the amount of time the TPM manual was studied. This showed that the amount of time the TPM manual was studied at least partially explained how motivation to study and study self-efficacy predicted TPM performance. In other words, the relationships were reduced after controlling for the amount of time the TPM was studied, but the direct effect for both motivation to study and study self-efficacy to TPM performance was still significantly different from zero.

About 24% of the variance was accounted for when motivation to study and study self-efficacy were entered together during the same step into a regression equation predicting total hours studying ($R^2 = .242, p < .001$). Similarly, when motivation to study and study self-efficacy were entered together during the same step into a regression equation predicting time studied during the final two weeks before the test was administered about 30% of the variance was accounted for ($R^2 = .303, p < .001$). Finally, about 18% of the variance was accounted for when motivation to study and study self-
efficacy were entered together during the same step into a regression equation predicting TPM test scores \((R^2 = .180, p < .001)\). This lent support to the concept that motivation to study and study self-efficacy played a role in the prediction of TPM test performance through the amount of time the TPM manual was studied.

*Black White Mean-Score Differences and Adverse Impact on the TPM Test as a Function of Study Time*

Research Question 2 asked if mean effect-size differences and adverse impact against Black test takers would decrease as the amount of study time reported by both Black and White test takers increased. First, it should be noted that TPM test scores generally increased as study time increased for both Black and White test takers (see Tables 12 and 13). Similarly, passing rates on the TPM test generally increased for both Black and White test takers as the amount of study time increased. Both of these findings were in line with the test of Hypotheses 3a, which had found a positive significant relationship between study time and TPM test performance.

Next, the Black-White mean effect-size differences in TPM test scores generally decreased as the total amount of study time increased (see Figure 7). However, the pattern of the Black-White mean effect size differences in TPM test scores based on the amount of time studied during the final two weeks before the test was administered was much less linear than anticipated (see Figure 8). Even so, the largest mean effect-size differences for both of the study-time interval conditions (i.e., total study time and the amount of time studied during the final two weeks before the test was administered) occurred for those who did not study the TPM manual at all and the lowest mean effect-
size differences occurred for those who studied the most for both total time studying and time studied during the final two weeks.

One rather surprising finding was that the mean effect-size differences on the TPM test and the measure of cognitive ability for the various study-time intervals were not more closely aligned, since cognitive ability is often considered the primary reason behind effect-size differences in scores between Black and White test takers. For example, the effect size difference on the measure of cognitive ability for those who reported not studying at all was more similar to the effect size difference on the measure of cognitive ability for those who studied the most than for any of the other study-time intervals, whereas the effect size difference in TPM test scores for those who reported not studying the TPM manual was the highest of all of the study-time intervals. This appears to suggest that cognitive ability may not have been a primary factor in explaining Black and White differences in TPM test performance for those who reported they had not studied the TPM manual. It is possible that other factors, such as previous firefighting experience or test wiseness may have played more important roles than cognitive ability for explaining this phenomenon for those who reported not studying the TPM manual. However, since those types of variables were not measured during the current study it was impossible to examine this proposition.

Even so, it should be noted that even though the Black-White effect-size differences on the TPM test were relatively small for those who reported not studying prior to the test being administered, the average TPM test scores of those job candidates were at least 20 raw-score points lower than those who had studied the most. Thus, the relatively large Black-White mean effect-size differences for those who did not study the
TPM manual had little practical significance in terms of successful test performance. However, this finding does appear to suggest that extraneous factors, such as previous firefighting experience, may have played a role in TPM test performance differences of Black and White job candidates.

Adverse impact against those who failed the TPM test was also examined in the testing of Research Question 2 (see Tables 16 and 17). It was found that adverse impact against Black test takers on the TPM test did not decrease steadily as the amount of study time increased for either total study time or time studied during the final two weeks before the test was administered. Instead, the pattern was non-linear, with the highest levels of adverse impact occurring for those who studied a total of 16 to 30 hours overall and those who studied 11 to 20 hours during the final two weeks before the test was administered. Even so, the lowest levels of adverse impact on the TPM test did occur when both Black and White test takers had studied the most overall and those who studied the most during the final two weeks prior to the test being administered. Somewhat unexpectedly, but in line with the proposition that increased study time would benefit Black test takers more than White test takers, the passing rate for those who studied the most during the final two weeks was higher for Black test takers than for White test takers.

Upon reflection it did not seem surprising that lower levels of adverse impact occurred when both Black and White test takers studied the TPM manual for relatively short periods of time as when compared to those who studied somewhat more. As mentioned previously, the average scores on the TPM test for those who reported that they had not studied the TPM manual were rather low when compared to those who had
reported they had studied the manual. Thus, it appeared logical that both Black and White test takers who had not studied the TPM manual or who had studied the TPM manual for only some minimal period of time would have performed quite poorly, probably at near chance levels.

The differences in the patterns found in the current study between mean effect-size differences and adverse impact revealed one of the reasons why researchers must be extremely careful when examining different indices of potential adverse impact (such as mean effect-size differences and Fisher-exact test findings) or attempting to explain why such differences may have occurred. It is extremely likely that a much less balanced conclusion about adverse impact would have been reached if the current research had relied on the findings from only subgroup mean differences.

In summary, Black-White effect-size differences in TPM test scores did generally decrease as study time increased. However, the trend was suggestive, but non-linear, for adverse impact on the TPM test when an unmodified Angoff cutoff score was used to determine who passed or failed the TPM test. These findings point out the importance of carefully interpreting the findings from different type of statistical analyses that are commonly used to determine potential adverse impact against protected groups of test takers. Although mean effect-size differences are extremely important for research purposes, adverse impact determinations based on Fisher Exact tests would appear to play a more important role for employers when they must determine significantly different hiring rates between subgroups as specified by the federal Uniform Guidelines on Employee Selection Procedures (1978).
Relationship Between Additional Study Techniques and TPM Test Performance

The analyses examining Research Question 3 indicated that study techniques, such as studying with a tutor or study group, setting aside a certain time to study, and the number of times a TPM practice test was taken added only about two percent or less variance to the prediction of TPM test scores above and beyond that of cognitive ability and time studying the TPM manual. This appeared to be of little practical significance. However, this should not necessarily be interpreted to mean that study techniques such as these would not be helpful for all job applicants preparing for TPM tests. Most likely, engaging in additional study techniques increased study time, so controlling for study time cancels this effect out. It this is true, these study techniques can potentially play an important role in the prediction of TPM performance.

Ryan, Ployhart, Greguras, and Schmit (1998) proposed that one reason for the finding of a very limited relationship between test preparation programs and test performance may be due to lower-ability performers being more likely to seek out and attend test preparation sessions. Furthermore, it is possible that these types of techniques might be more helpful and appropriate for some test takers than others. For example, test takers who are poorly motivated to study and succeed would likely not benefit from these techniques whereas they might better benefit those who were highly motivated. Research examining how individual differences could influence the effectiveness of different study strategies is encouraged.

Major Findings

There were three major findings from the current study that bear repeating. The first was the lack of significant Black-White differences in the levels of motivation to
study and study self-efficacy. This is especially important since previous studies have reported that Black job applicants typically report possessing less test-taking motivation than White job applicants. The second was that Black test takers reported studying significantly longer than White test takers, although in practical terms this difference was small, it does indicate that Black test takers did not study less than White test takers. Both of these findings appeared to provide strong support for the “natural” interaction hypothesis (Asher and Sciarrino, 1974; Smith 1991) for the testing of job applicants. This leads to the third major finding, which was that offering a test that allowed for the natural interaction of ability, motivational beliefs, and study strategies, resulted in lower subgroup mean-score differences and adverse impact than a measure of ability.

Limitations

This study does have limitations that must be considered. First, the current study examined the TPM testing process as it was used during a single test administration, which may have possessed unique characteristics. For example, the $d$ between Blacks and Whites on the ability tests used for firefighter selection during the study by Ryan, Ployhart, Greguras, and Schmit (1998) was about .96, whereas the Black-White $d$ on the cognitive ability measure in the current study was .78. However, some of the reduced $d$ on the measure of cognitive ability in the current study could have been due to a relative ceiling effect for scores for White test takers on that measure.

The reduced Black-White $d$ also could have been at least partially due to the requirement that job applicants possess a basic Emergency Medical Technician (EMT) certification prior to a job offer being made, which appears to be somewhat unique to the employers involved in the current study. Basic EMT certification in the jurisdiction
where the test was administered included about 100 hours of instruction, plus ten hours of clinical practice (Rural Emergency Medical Services Outreach Project and University Center for Economic Development, 2004). Those seeking basic EMT certification were also required to pass a standardized skills test and a written multiple-choice examination. The requirement for basic EMT certification likely reduced the number of job applicants in the current study who lacked the cognitive ability to pass the certification process.

Regional differences may also have played a role in the lower Black-White score differences on the cognitive ability test in the current study. For example, the Ryan et al. study examined entry-level firefighter testing offered by a mid-western city whereas the current study examined job seekers who were testing for employment as firefighters for a city in the far-western United States. It is possible that job applicants from different parts of the country might have experienced different educational opportunities and other factors when they were growing up. With that in mind, future researchers should consider differences that might potentially be due to the geographic location of their studies when attempting to generalize to other locations.

The highest level of reported education was not significantly different between Black and White test takers in the current study (see Table 4). In addition, analyses revealed there were no significant differences in the current study between Black ($M = 0.200, SD = 0.401$) and White ($M = 0.147, SD = 0.355$) participants as to whether they had attended college or not, $t(179.453) = 1.532, p = 0.127$. Conversely, Black test takers were less likely to have attended college than Whites in the Ryan et al. (1998) entry-level firefighter study ($\chi^2 (1) = 35.33, p < .001$).
All of these differences taken together suggested that the participants in the current study might have been in some ways different from participants of other studies, such as those who were included in the firefighter pre-employment testing research conducted by Ryan et al. However, significant individual differences between groups are often found in different studies and those differences do not necessarily indicate that the results from those studies are not valid or useful.

Second, it has been well known for many years that increasing the length of a measure increases its consistency (e.g., Thurstone, 1931; Adkins, Bridges, Forer, McAdoo, & Primoff, 1947; Cuerton, 1965). This positive relationship between the length of a measure and its reliability is typically shown using the Spearman-Brown prophecy formula (Cascio, 1998). However, during the current study only one survey item was used to measure each study strategy and motivational characteristic of interest, resulting in less reliable measures of the characteristics than if a greater number of items had been used. It is possible that the relationships might have played out differently if more reliable measures had been used.

A third limitation was the use of archival data from a measure that employed non-equivalent interval-style scales when asking respondents to indicate the amount of time they spent studying. This was especially problematic for the longest study-time interval since those who studied 21 hours during the final two weeks were treated as the same as those who may have studied many times longer.

A fourth limitation was the reliance on a self-report device that relied totally on the memory of the participants for indicating the number of hours they studied. A detailed examination of the potential limitations of using self-report data was previously
discussed during Chapter 3 and will not be repeated here. However, Fowler (1993) indicated that research has consistently shown that when participants in a study do not accurately recall a specific amount in response to questions they tend to estimate the answers based on their usual patterns rather than on their actual behavior for the situation in question. This suggests that participants in the current study who did not accurately recall the number of hours they studied might have substituted the number of hours they typically would have studied instead, which might have confounded the results. Unfortunately, the archival data used during the current study was not sufficient to allow such a determination to be made. A suggestion on how to address this limitation is presented in the next section of this chapter.

Next, the Black-White mean effect-size difference in scores on the TPM test for those who reported not reading the TPM manual at all was $d = 1.104$ and for those who reported not studying during the final two weeks before the test was administered it was $d = 1.151$. This suggested that Whites who did not study the TPM manual were much more likely to correctly respond to the TPM test items than Blacks who did not study the TPM manual. In 1997 Barrett, Miguel, and Doverspike suggested, “if one’s aim or intent is to reduce race differences, than it would seem to be a huge disadvantage to start out with a large mean difference based only on the questions without passages” (page 23).

These relatively large mean-score differences on the TPM test for those who reported not reading the associated passages were especially troubling since the average mean-score difference for the measure of cognitive ability for those who reported not studying the TPM manual at all was only $d = 0.540$ and for those who reported not
studying the TPM manual during the final two weeks before the test was administered was $d = 0.536$.

There appear to be at several potential explanations for the Black-White TPM test score mean differences for those who reported not studying the TPM manual. First, it appears possible that some test takers might have been more “test wise” than others, which could have contributed to the test-score differences. Test wiseness refers to a person’s ability to use situational and test characteristics to improve test scores (Millman, Bishop, & Ebel, 1965). Guion (1998) lamented that there has been a lot written about test wiseness, but that is rarely systematically studied.

One study that did examine test wiseness was reported by Flippo and Caverly (2000). They found that test wiseness did play an important role in how test takers performed during testing and they encouraged the teaching of test wiseness to all test takers, which they felt should increase the validity of the testing process. McKay and Doverspike (2001) suggested that test wiseness training should further include an emphasis of the value of studying, practice, and hard work; encourage a positive attitude toward the testing process; and attempt to increase test-taking motivation. Finally, Barrett, Miguel, and Doverspike (1997) reported that some of the success test takers had responding to public safety reading test questions without being presented the associated reading passages was a function of cognitive ability.

However, Powers and Leung (1995), questioned whether test wiseness was the primary factor for success when test takers were asked to respond to test questions without being presented the associated passages. Powers and Leung (1995) examined whether verbal skills played a more important role than test wiseness when test takers
were presented SAT reading comprehension test questions without the accompanying reading passages. They found that the strategies used by those who did better answering these questions reflected the use of verbal reasoning more than test-wiseness skills. More specifically, they found that the test takers who attended to the consistencies in the questions and used this information to reconstruct the theme of the passages, which had not been presented to them, did better than those relying on test wise alone.

There are other potential explanations as to why White test takers who reported not studying the TPM manual performed better on the TPM test than Black test takers who reported not studying. It is possible that White test takers had greater previous test taking experience, and possibly, greater experience with TPM type tests.

Catron and Thompson (1979) conducted studies of gains on the Wechsler Adult Intelligence Scale (WAIS) test scores. They found that the experience of taking a test can positively affect the results on any similar test afterward. Furthermore, Burke (1997) examined the effects of retesting on aptitude tests given to Royal Air Force applicants. He found that retesting resulted in statistically significant gains in test scores, without regard to the amount of time between the first test and retest administrations.

Another potential explanation for White test takers performing better than Black test takers on the TPM test for those who reported they had not studied the TPM manual might have been related to the amount of previous firefighter experience the test takers may have possessed. As indicated by Wollack in 1994, Blacks typically score lower on firefighter civil service entry-level exams than Whites, which results in fewer Blacks being hired. Test takers with previous experience would appear to have had a competitive edge over those without experience since the TPM manual was specifically designed to
include job-related knowledge about firefighting. Unfortunately, information related to
the test takers’ prior experience in the fire-related fields was unavailable for the current
study, so it was impossible to determine whether White test takers actually did have more
firefighting experience than Blacks. However, it would appear that a very cautiously-
drawn inference could be made about the opportunity Black test takers from Southern
California might have had for previous firefighter job experience in general based upon
data from the 2000 Census.

A review of the zip codes of all test takers who took the tests examined in the
current study revealed that the vast majority were from Southern California. U. S. Census
(2000) data indicated that Blacks were underrepresented as firefighters when compared to
the total available workforce in seven of the ten counties traditionally identified as being
located in Southern California (Imperial, Kern, Los Angeles, Orange, Riverside, San
Bernardino, San Diego, San Luis Obispo, Santa Barbara, and Ventura). Conversely, a
greater percentage of Whites self-identified as firefighters in all ten of those same
Southern California counties when compared to the available workforce. Furthermore,
zero percent of Blacks in the Census study self-identified as firefighters in three of the ten
counties (Imperial, San Luis Obispo, and Santa Barbara).

Of course, this does not necessarily mean that Black test takers had less
firefighting experience than White test takers in the current study. It does, however,
indicate that Blacks, in general, would have had less of an opportunity to gain experience
as firefighters than Whites in these ten Southern California counties, which could have
played at least a partial role in the test-score differences on the TPM test for those who
reported not studying the TPM manual.
Suggestions for Future Research

First, future researchers examining the influence of study time on TPM test scores are encouraged to use techniques that would allow participants to more accurately indicate the specific amount of time studied. For example, participants could be asked to provide the actual number of hours they studied rather than selecting from one of five intervals.

Along similar lines, future researchers might consider using methodologies that would help participants more accurately identify the amount of time studying. This could include the use of diaries to assist test takers in keeping track of the time they spent studying. Bolger, Davis, & Rafaeli (2003) indicated that diaries can help with obtaining accurate within-person information.

Next, additional studies of the relationship between TPM test scores and job and/or academy performance are encouraged to determine whether TPM-style testing possesses sufficient validity should be considered as an alternative selection device under Section 3B of the federal Guidelines. Section 3B states if two or more selection devices equally serve an employer’s legitimate interest in efficient and trustworthy workmanship, and which are substantially equally valid for a given purpose, the device with lower adverse impact should be used. The reduced adverse impact shown in the study might be moot if this form of testing fails to possess a sufficiently high level of validity.

Another area that should be explored is whether the advantage shown for Blacks who studied longer during the current study might equally benefit them if additional time was provided during a traditional reading comprehension test as suggested by Walczyk and colleagues (Walczyk, Kelly, Meche, & Braud, 1999; Walczyk, Marsiglia, Bryan, &
Naquin, 2001; Walczyk & Taylor, 1996). Future researchers might wish to offer traditional-style reading comprehension test questions under lenient time-restriction conditions along with a TPM-style test and then compare the adverse impact differences in order to determine if TPM-style testing uniquely provides additional benefits, and/or decreases the level of adverse impact, when compared to more traditional styles of tests of reading and learning.

Next, it would appear that a greater proportion of the test takers would benefit from taking sample test items if they were made available to all test takers without charge. Future researchers may wish to determine whether the cost of the practice test questions prevented some of the test takers from taking advantage of this preparation tool, which may have affected potential differences in TPM test scores.

Finally, it should be noted that just because job seekers with lower ability levels might be able to compensate when allowed additional study time is allowed does not mean that they will compensate. Future studies should attempt to identify the circumstances under which low ability test takers are more likely to compensate, and the variables which might affect compensation.

It appears that many of the relationships in the current study could potentially be better understood through of the use of laboratory studies. Laboratory studies allow researchers to more accurately measure individual difference variables such as the amount of time studying and the types of study strategies that are used. Researchers can also manipulate these and other variables in the lab in ways that could not be accomplished in the field.
Implications for Professional Practice

The evidence from the current study found that TPM-style testing had lower adverse impact against Black job seekers for the position of firefighter than a more traditional measure of cognitive ability given during the same test administration. Furthermore, when variables that were also significantly related to TPM test scores were controlled, Black test takers benefited more from increased study time than White test takers. With that in mind, organizations wishing to reduce adverse impact when the ability to read and study information is needed for job success might consider using TPM-style testing rather than more traditional tests of cognitive ability. Of course, it would make sense to substitute more traditional measures of cognitive ability with TPM-style testing only when the measures are substantially equally valid, or when the TPM-style test is more valid than the measure of cognitive ability, for selection purposes.

Next, if TPM-style selection devices are used they should be designed to minimize a test taker’s ability to correctly respond to the test questions without having read the associated passages. This could be accomplished by pilot-testing the test questions on a group of test takers without the associated passages being presented and removing those items which could be correctly answered based on test wiseness, previous test experience, or previous job experience.

However, it would also be helpful to determine whether minimizing the ability of test takers to correctly respond to test questions without having read the associated passages would adversely affect the test’s criterion-related validity. This could be accomplished by comparing the criterion-related validity of those who study the TPM manual prior to testing to those who do not read the manual. If the criterion-related
validity between TPM test scores and some job-related performance criteria is the same or stronger for those who had not read the manual as it is for those who did study the manual, then the utility of the TPM-testing process would appear to be called into serious question.

Finally, one concern previously expressed about using alternative selection measures has been that attempting to reduce adverse impact against one group of test takers might inadvertently increase adverse impact against another protected group. For example, Ryan, Ployhart, and Friedel (1998) cautioned that making changes in a testing process to minimize adverse impact against minority groups might introduce adverse impact against women. This would be especially troublesome because women are typically under-represented in the firefighting profession (Rosell, Miller, & Barber, 1995). Fortunately, even though TPM testing reduced Black-White score differences, the test-score differences between men and women were not significantly different. This should ease a major concern that some practitioners might have had about using this approach to testing.
CHAPTER VI

SUMMARY

Human resource professionals have been earnestly searching for effective pre-employment selection measures of the ability to perform job-related tasks, which would result in lower levels of adverse impact against minority-group job candidates than traditional measures of pure cognitive ability (Viswesvaran & Ones, 2002). Some researchers have suggested that trainability testing could provide one such alternative (e.g., Aramburu-Zabala & Riera, 2004; Campbell, 1982; Higuera and Riera, 2004; Reilly and Israelski, 1988; Robertson & Downs, 1989). One style of trainability testing that has been designed to measure the cognitively-loaded tasks of reading and learning information, while still reducing potential adverse impact, is TPM testing.

During the current study the TPM testing process consisted of job applicants for entry-level firefighter positions being instructed to study a manual containing job-related information and then, during the testing session, answer questions from memory about information contained in that manual. This process was designed to mimic the type of reading and learning that newly-hired firefighters must perform while attending a typical fire academy and which more experienced firefighters must perform on the job.

In the current study, it was found that the standardized test-score differences for the TPM test were significantly lower for the TPM test ($d = .334$) than for a measure of cognitive ability ($d = .764$), which was administered during the same session.
Furthermore, adverse impact against Black test takers on the TPM test was 2.394 standard deviations, whereas adverse impact against Black test takers on the measure of cognitive ability was 7.195 standard deviations. Thus, the TPM test did result in a smaller difference between subgroups and lower adverse impact. In an attempt to discover why this effect occurred, a simple and complex model of TPM test performance were developed and tested with 2,078 firefighter applicants for an entry level position.

In support of the model, the total amount of time a TPM manual was studied ($r = .48, p < .001$), the amount of time the TPM was studied during the final two weeks before the test was administered ($r = .56, p < .001$), and cognitive ability ($r = .38, p < .001$) were significantly correlated with TPM test performance. Unlike previous studies that had found the relationship between study time and performance to be somewhat equivocal, the effect size was relatively strong in the current instance. Furthermore, mean score differences on the TPM test generally were reduced as the amount of time studying increased for both Black and White test takers.

However, a hypothesized statistical interaction between cognitive ability and study time was not found, at least not in the predicted direction. In the current study those test takers who performed better on the measure of cognitive ability benefited more from additional study time than those who performed worse on the measure of cognitive ability, although the size of this effect was very small from a practical perspective. This was disappointing since early versions of the Compensatory-Encoding Model (C-EM; Walczyk, 1993, 2000) had originally postulated that allowing additional time to study would more likely benefit those with lower ability levels more than those with higher ability levels.
Instead, this study revealed that the amount of time studying the TPM manual was the primary reason for the reduction in Black-White mean score differences on the TPM test as compared to differences on a measure of cognitive ability. This showed the wisdom behind the “interaction hypothesis” of work-sample testing put forth by Asher and Sciarrino (1974), which was later endorsed by Smith (1991). This hypothesis proposed that work-sample testing, which allows for a test’s various components to combine “naturally,” is likely to better predict performance of complex tasks than an additive combination of scores from measures of discrete abilities; especially since additive combinations of scores would overlook aptitudes that interact.

The effect of race on TPM test performance was fully mediated by cognitive ability. This is in line with the preponderance of previous research that Black-White mean score differences on tests are typically explained by differences in cognitive ability or $g$.

Black and White test takers in the current study indicated similar levels of motivation to study and study self-efficacy. This is counter to other studies of pre-employment testing (e.g., Ryan, Ployhart, & Friedel, 1998), which have generally reported Black test takers indicating lower levels of motivation and self-efficacy than White test takers.

Furthermore, motivation to study and study self-efficacy added no practically significant variance to the prediction of TPM test performance above and beyond a combination of time studying the TPM manual and cognitive ability. However, both motivation to study and study self-efficacy added statistically and practically significant
variance to the prediction of the amount of time studying, thereby having an indirect effect on TPM test performance.

Thus, the TPM test examined in the current study resulted in significantly lower subgroup mean-score differences, and in lower levels of adverse impact, than a measure of cognitive ability that was administered at the same time. The primary cause for this lower adverse impact was a function of the amount of time the TPM manual was studied. Furthermore, both Black and White test takers reported similar levels of motivation and study self-efficacy in the current study, whereas many previous studies have found that White test takers generally have higher levels of both of these individual difference variables during pre-employment testing.

Of course this study was not without limitations, including the self-report method by which the amount of time studying was measured and the use of single-item scales for measuring many of the various individual difference variables. Hopefully, the promising results will encourage other researchers to more closely examine the use of TPM-style testing as an alternative method for selecting employees when its use is justified by a thorough analysis of the job.


158


160


Holziner, K. J. (1936). Recent research on unitary traits. *Character and Personality, 4*, 335-343.


National Institute of Child Health and Human Development. (2000). Report of the National Reading Panel. Teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction:


APPENDICES
APPENDIX A

APPROVAL LETTER FROM INSTITUTIONAL REVIEW BOARD

Office of Research Services and Sponsored Programs
Akron, OH 44325-2102
(330) 972-7666 Office
(330) 972-6781 Fax

April 5, 2004

James Kuthy
2236 Woodside Lane #4
Sacramento, California 95825

Mr. Kuthy:

The University of Akron’s Institutional Review Board for the Protection of Human Subjects (IRB) completed a review of the protocol entitled “Examination of the Relationship of General Mental Ability and Study Time to Scores on a Study Guide Style Reading Test used for Pre Employment Testing”. The IRB application number assigned to this project is 20040402.

The protocol qualified for exemption from continuing IRB review on April 2, 2004. The protocol represented minimal risk to subjects. Additionally, the protocol matched the following federal category for exemption:

- Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.

If you propose changes to this protocol, an Application for Continuing Review Form must be completed and submitted to the Office of Research Services.

Please retain this letter for your files. If the research is being conducted for a master's thesis or doctoral dissertation, the student must file a copy of this letter with the thesis or dissertation.

Sincerely,

[Signature]
Sharon McWhorter, Associate Director

Cc: Linda Subich, Department Chair
    Dennis Doverspike, Advisor
    Rosalie Hall, IRB Vice Chair
    Phil Allen, IRB Chair

The University of Akron is an Equal Education and Employment Institution
APPENDIX B

CHARACTERISTICS SURVEY QUESTIONS

Please answer the following 15 questions as honestly as possible. Your answers to the following questions will NOT count towards your final score. If you are NOT going to answer a question honestly, please skip that question. The [Redacted] or [Redacted] Fire Department personnel will NOT be told of your answers to these questions. These questions are designed for research purposes only to help us improve the testing process. Thank you.

The following questions pertain to ONLY the Test Preparation Manual portion of today’s test.

1. The highest level of education I have completed is:
   A. a High School Diploma
   B. a G.E.D.
   C. an Associate’s Degree or Vocational School Certification
   D. a Bachelor’s Degree
   E. above a Bachelor’s Degree

2. My average overall grade in High School was:
   A. A
   B. B
   C. C
   D. D
   E. Less than D

3. I spent a total of about _________ hours studying the Test Preparation Manual (TPM) with a tutor or a study group. (For this question ONLY, do not include any time you spent studying alone.)
   A. no (I never read or studied the Test Preparation Manual with a tutor or study group.)
   B. 1 to 15
   C. 16 to 30
   D. 31 to 45
   E. more than 46
4. An optional Test Preparation Manual (TPM) practice test is offered to job applicants both on the Internet and in printed form. I took the Test Preparation Manual (TPM) practice test:

A. not once. (I never took a Test Preparation Manual practice test.)
B. one time.
C. two times.
D. three times.
E. four or more times.

5. I completely read the Test Preparation Manual (TPM) in order to prepare for today’s test:

A. not once. (I never completely read the Test Preparation Manual.)
B. one time.
C. two times.
D. three times.
E. four or more times.

6. The Test Preparation Manual (TPM) was very easy for me to read and understand.

A. Strongly disagree
B. Disagree
C. Neither agree nor disagree
D. Agree
E. Strongly agree

7. I felt extremely motivated to study the Test Preparation Manual (TPM) to prepare for today’s test.

A. Strongly disagree
B. Disagree
C. Neither agree nor disagree
D. Agree
E. Strongly agree
8. I had a **quiet place to sit down and study** the Test Preparation Manual (TPM) when preparing for today’s test.

   A. Strongly disagree
   B. Disagree
   C. Neither agree nor disagree
   D. Agree
   E. Strongly agree

9. I spent a **total** of about ______ hours reading/studying the Test Preparation Manual (TPM) **since I received it**.

   A. no (I never read or studied the Test Preparation Manual)
   B. 1 to 15
   C. 16 to 30
   D. 31 to 45
   E. more than 46

10. **During the past two (2) weeks** I spent a total of about ________ hours reading/studying the Test Preparation Manual (TPM).

    A. no (I did not study or read the Test Preparation Manual in the past two weeks.)
    B. 1 to 5
    C. 6 to 10
    D. 11 to 20
    E. 21 or more

11. When preparing for today’s test, I believed that the more I studied the Test Preparation Manual (TPM) the **higher I would score** on today’s test.

    A. Strongly disagree
    B. Disagree
    C. Neither agree nor disagree
    D. Agree
    E. Strongly agree

12. Being hired by the [Redacted] or [Redacted] fire departments as a firefighter is **extremely important** to me.

    A. Strongly disagree
    B. Disagree
    C. Neither agree nor disagree
    D. Agree
    E. Strongly agree
13. I am **smart enough** to be a firefighter.

   A. Strongly disagree  
   B. Disagree  
   C. Neither agree nor disagree  
   D. Agree  
   E. Strongly agree

14. I set aside **certain times to study** the Test Preparation Manual (TPM) for today’s test.

   A. Strongly disagree  
   B. Disagree  
   C. Neither agree nor disagree  
   D. Agree  
   E. Strongly agree

15. The average newly-hired firefighter in training **reads written text** approximately _____ hour(s) each day he or she is attending fire academy classes for the [Redacted] or North [Redacted] Fire Departments.

   A. less than one (1)  
   B. one (1)  
   C. two (2)  
   D. three (3)  
   E. four (4) or more

Thank you very much for honestly answering the previous 15 questions. As mentioned previously, your responses to these questions will NOT count towards your final grade, and **no** one from the fire department will be told of your individual responses on this section.
APPENDIX C

TPM MANUAL SAMPLE TEXT

WATER VACUUM

The water vacuum is a device that is carried on the operator’s back and has a hand-held vacuum hose and nozzle. It is used to suck up water after firefighting or at a spill. It must be plugged into an AC electrical source in order to operate. It has the ability to pick up approximately 3-1/2 gallons in just a few minutes.

OXY-ACETYLENE CUTTING TORCH

The cutting torch is used to cut various thickness of metal for forcible entry and rescue. Unlike saws or shears, a cutting torch can sever extremely thick, heavy metal.

A firefighter must use caution when operating an oxy-acetylene torch. It can be dangerous because of its high heat flame and potential for igniting flammable gases.

There are two tanks of compressed gas that make this torch work, oxygen and acetylene. These are connected to the torch by supply hoses. Acetylene has one of the highest flammable ranges. It will burn in an atmosphere of very little oxygen or in one with a great deal of oxygen.

COMMON CODE VIOLATIONS

Exits – The location, marking, size, hardware, number, and maintenance of exits are detailed in the building code. Exits are of particular concern, for they may be the only means of escape for building occupants. For example, panic hardware is a bar which extends across at least one-half the width of each door leaf, which will open if subjected to pressure from the inside. This means that trapped victims can always open the door going out. The width of every stairway shall be maintained and trim shall not reduce the required width by more than 3-1/2 inches.

APPENDIX D
SAMPLE TPM TEST QUESTIONS

Fallout shelter signs are colored _________________.

A. yellow and black
B. red and black
C. red and yellow
D. green and yellow

Gasoline, kerosene, and jet fuels all have vapors that are ___________ times heavier than air.

A. ½ to 1
B. 1 to 2
C. 1 ½ to 2 ½
D. 2 to 3

Upon hearing the emergency evacuation signal, all personnel should _________________.

A. report to the front of the fire engine
B. locate their partner
C. report to the back of the fire engine
D. locate their supervisor