TEST-WISENESS TRAINING: AN INVESTIGATION
OF THE IMPACT OF TEST-WISENESS IN AN EMPLOYMENT SETTING

A Dissertation
Presented to
The Graduate Faculty of The University of Akron

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Susan Elizabeth Houston
December, 2005
ABSTRACT

The current study examined ethnic group differences in test-wiseness and the extent to which test-wiseness training may eliminate these differences in a sample of 87 firefighters from three different metropolitan areas. As part of a larger eight hour training program on assessment centers, subjects were given two measures to assess their level of test-wiseness (learning and behavior pre-tests). Subjects were then instructed on test-wise strategies involving item construction. Following this training, subjects were given a measure to assess their reactions to the training program as well as two post-test measures (learning and behavior).

The current research revealed that there were no significant differences between whites and African Americans on the pre-test Learning measure and the pre-test Behavior measure. While overall, training had a positive impact on subjects' abilities to identify the test-wiseness cues on the Learning measure with subjects showing a significant improvement, subjects showed only marginal improvements on the Behavior measure. In addition, rather than diminishing group differences, test-wiseness training appeared to have no significant race by training effect on the Learning measure and appeared to exacerbate the differences between whites and African Americans on the Behavior measure.
ACKNOWLEDGEMENTS

I can’t believe how long I have looked forward to finally being able to put this dissertation behind me. There are so many people whose support and friendship have been invaluable in helping me get to the point where I no longer have to worry about this thing hanging over my head.

First of all I have to thank Allen. He has been by my side through the whole long and stressful process. He is my best friend, biggest supporter, anxiety reliever, reality checker and love of my life. In many ways my two amazing sons, Benjamin and Adam were also instrumental in me finally getting this done. Having the luxury of being home with Benjamin and the impending birth of Adam really put me in the right place to put all of the pieces together. I am also so lucky to have such wonderful parents, Joe and Jane, and family (John, Barbara, Chloe, Joel, Sherry, Christopher, Meredith) who have supported me in every way. I also have to thank Allen’s family (Bernie, Elaine, Syd, Fern, Charles, Alana, Samantha, Lowell, Jennifer, Jack and Isabel) for never giving up hope that I would someday finish even though my estimates were continually getting longer and longer.

There are also many members of the Akron faculty that were incredibly helpful. In particular I would like to thank Paul Levy for being so supportive and
helpful in getting me through all of the final hurdles, Dr. Barrett for teaching me to think like an I/O psychologist, Rosalie Hall for her patience and guidance during an incredibly busy and stressful time, and Mike McDaniel and Dennis Doverspike for their flexibility and humor.

I would also like to thank Kasey Weidman for editing and compiling all of the materials to satisfy all of the formatting requirements, a headache I was dreading. Finally, I would like to thank all of the friends who helped me get through. Anthony Mellinger, Dave Bernal, Dave Snyder, LaRae Jome, Joelle Elicker, Elizabeth Rychcik, Elaine Engle, John Johnson, Greg Reid, Ted Axton, Cathy Callahan, Earl Hartman, Diane Govern, and all the others I didn’t specifically mention. You all made even the toughest times bearable and your sense of humor helped make the years in Akron a time I will always value.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ix</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LIST OF FIGURES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

## CHAPTER

I. STATEMENT OF THE PROBLEM ................. 1

II. LITERATURE REVIEW .......................... 5

  - Background ................................ 6
  - Is Test-wiseness Research Generalizable to Employment Settings? .................. 11
  - Can Test-wiseness be Effectively Trained in an Employment Setting? ............ 12
  - Group Differences and Test-wiseness ............................................... 14
  - Overview and Hypotheses ............................................................... 17

III. METHODOLOGY ................................ 20

  - Subjects .................................. 20
  - Procedure ................................ 22
  - Measures .................................. 23
  - Scale Development ......................... 29
  - Training Session ......................... 35
IV. RESULTS ................................................................. 37
  Reaction Results .................................................. 37
  Learning Results .................................................. 40
  Behavior Results .................................................. 43
  Exploratory Analyses ............................................. 49
  Student Data as a Non-Equivalent Control Group .......... 56
V. DISCUSSION ............................................................. 58
  Does Test-wiseness Generalize to Employment Settings? . 58
  Can Test-wiseness be Trained? ................................. 59
  Group Differences in Test-wiseness ......................... 62
  Does Training Alleviate Group Differences in Test-wiseness? 63
  Theoretical Explanations for Findings ....................... 64
  Limitations and Methodological Explanations for Results . 70
  Implications ......................................................... 71
REFERENCES ................................................................. 74
APPENDICES ................................................................. 81
  A. Human Subjects Approval ................................. 82
  B. Biographical Information Sheet .......................... 83
  C. Learning Measure ............................................. 84
  D. Behavior Measures .......................................... 86
  E. Training Guide ................................................. 96
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Means and standard deviations for demographic variables</td>
<td>21</td>
</tr>
<tr>
<td>2.</td>
<td>Overview of cues used in behavior measures</td>
<td>25</td>
</tr>
<tr>
<td>3.</td>
<td>Pilot study p values for pre-test items counterbalanced for order effects</td>
<td>32</td>
</tr>
<tr>
<td>4.</td>
<td>Pilot study p values for post-test items counterbalanced for order effects</td>
<td>33</td>
</tr>
<tr>
<td>5.</td>
<td>Pilot study p values for pre-test and post-test Items by test-wiseness cue</td>
<td>35</td>
</tr>
<tr>
<td>6.</td>
<td>Means and standard deviations for the reaction measure by race</td>
<td>38</td>
</tr>
<tr>
<td>7.</td>
<td>Descriptive statistics and intercorrelations among variables used in study</td>
<td>39</td>
</tr>
<tr>
<td>8.</td>
<td>Repeated measures effect sizes for pre-test and post-test measures for overall sample and broken apart by ethnic group</td>
<td>41</td>
</tr>
<tr>
<td>9.</td>
<td>Means and standard deviations of learning measure by race</td>
<td>42</td>
</tr>
<tr>
<td>10.</td>
<td>Means and standard deviations for behavior measure by race</td>
<td>44</td>
</tr>
<tr>
<td>11.</td>
<td>Means and standard deviations of cue dimensions for behavior pre-test and post-test by race</td>
<td>46</td>
</tr>
<tr>
<td>12.</td>
<td>Intercorrelations of cue dimensions on the behavior measures</td>
<td>47</td>
</tr>
<tr>
<td>13.</td>
<td>Means and standard deviations of learning and behavior measures by age group</td>
<td>50</td>
</tr>
<tr>
<td>14.</td>
<td>Means and standard deviations of job knowledge test scores</td>
<td>51</td>
</tr>
</tbody>
</table>
15. Descriptive statistics and intercorrelations of job knowledge test scores and test-wiseness and demographic variables .............. 53
16. Correlations between job knowledge test scores, test-wiseness variables, and demographic variables for whites and African Americans .................................................. 54
17. Correlations between job knowledge test scores and test-wiseness cues for whites and African Americans .............................. 55
18. Means and standard deviations for behavior measures for race controlled by age and job knowledge test scores .................... 55
19. Means and standard deviations of behavior measures by students and employees ................................................................. 57
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hypothesized results for Learning and Behavior measures by ethnic group</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>Overview of measures used in method</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>Results on pre- and post-test learning and behavior measures by ethnic groups</td>
<td>42</td>
</tr>
<tr>
<td>4</td>
<td>Results on pre- and post-test learning and behavior measures by age groups</td>
<td>50</td>
</tr>
</tbody>
</table>
CHAPTER I

STATEMENT OF THE PROBLEM

Every year, organizations test thousands of individuals for selection or promotional purposes (Riggio, 1996), with paper and pencil tests being the most commonly used type of test in employment settings (Muchinsky, 1987). A major concern that organizations must consider is the adverse impact that many of these tests have and the potential legal action that may occur. Therefore, industrial/organizational psychology is continually involved in research to investigate strategies which may prove to reduce adverse impact (Barrett, Doverspike, Cellar, & Johnson, 1991).

In previous discrimination cases, one issue that has been raised is whether test-wiseness contributes to the difference between groups on employment test scores. Surprisingly, however, little research has actually investigated the impact of test-wiseness in selection or promotional settings. Instead, the research that does exist has predominantly involved educational settings and has used students as subjects. Therefore, the extent to which the test-wiseness literature generalizes to employment settings is an area that deserves greater attention given the potential motivational differences that may exist between students and employees (Arvey, Strickland, Drauden, & Martin,
1990; Jennings, 1953; Latham & Dossett, 1978). Therefore, the current research seeks to investigate whether ethnic group differences in test-wiseness exist and whether test-wiseness can be effectively trained when the testing is for the purpose of making a hiring or promotion decision. In addition, this research seeks to explore whether training can effectively eliminate or reduce ethnic group differences in test-wiseness.

While there has been little research on test-wiseness in the industrial/organizational psychology literature, the issue has received considerable attention in the education literature. Test-wiseness has widely been defined as an individual's ability to improve his or her test score by recognizing and utilizing cues in the test items, format or testing situation. This ability to improve scores is independent of an individual’s knowledge of the material on the test. Therefore, it is not surprising that education researchers have focused on test-wiseness given the potential impact it may have on students’ test performance. The most pertinent of this research to the area of employment testing and the legal defensibility of tests is research which has found evidence that test-wiseness may be a source of additional variance in test scores and a factor which may lower test validity (Dolly & Vick, 1986; Fagley, 1987; Sarnacki, 1979; Thorndike, 1951). If individuals are able to correctly answer questions without actual knowledge of the material, this would detract from the predictive utility of the test.

The concept of test-wiseness has been thought of as made up of various different aspects. For example, time management skills, attempting to answer all
questions, and fully understanding the directions of the tests are all considered to
be facets of test-wiseness (Millman, Bishop, & Ebel, 1965). Additional
components of test-wiseness have been defined as the characteristics of the
items themselves and individuals’ abilities to determine the correct answer
through item cues. Items can include cues such as the use of information from
other items, alliterative associations, longer correct alternatives, more specific
correct alternatives, etc. (Sarnacki, 1979). While there are various components
of test-wiseness that are of interest in employment and/or promotional settings,
this research focuses directly on embedded cues which may help individuals
decipher the correct alternative without relying on knowledge or experience with
the content area.

Within the test-wiseness literature, there are several questions that have
been debated and are still largely unanswered. This research seeks to explore
some of these issues. One such debate has involved whether test-wiseness can
be effectively trained. Another debate is whether there are ethnic group
differences in levels of test-wiseness.

In the educational literature, several researchers have found evidence of
the usefulness and effectiveness of test-wiseness training (Dolly & Vick, 1986;
Dolly & Williams, 1985). However, other researchers have argued that the
findings have been inconclusive regarding whether test-wiseness is a skill that
can be effectively trained (Scruggs & Lifson, 1985; Scruggs, White, & Bennion,
1986).
The research has also been mixed about whether ethnic group differences in test-wiseness exist. Several studies have reported evidence of ethnic group differences (e.g., Barrett, Miguel, & Doverspike, 1997; Diamond, Ayres, Fishman, & Green, 1976; Dreisbach & Keogh, 1982; Ebel, 1968; Kalechstein, Hocevar, & Kalechstein, 1988; Miguel, 1997). However, others have argued that evidence which has been used to support the idea that minorities are lacking in test-wiseness is often unconvincing and are often methodologically flawed (McPhail, 1984; Scruggs & Lifson, 1985). In addition, other studies examining the relationship of ethnic group and test-wiseness have failed to reveal significant findings (e.g., Benson, Urman, & Hocevar, 1986; Diamond, Ayres, Fishman, & Green, 1976; Yearby, 1975).

Given the research reviewed above which has reported the beneficial effects of test-wiseness training, it is quite possible that individuals who are lower in test-wiseness may be at a disadvantage in a testing environment (Crehen, Koehler, & Slatker, 1974; Moore, Schultz, & Baker, 1966). In particular, if there are ethnic group differences in test-wiseness such that African Americans tend to score lower in test-wiseness than whites, African Americans will be at a disadvantage in a testing environment. In other words, ethnic group differences in test-wiseness could mean that individuals in certain groups would be better able to answer questions without relying on prior knowledge or experience. This could have potentially important consequences in the selection and promotion of individuals.
One of the greatest concerns of organizations is selecting and promoting individuals who are competent and able to perform well. Consequently, the means by which organizations select individuals is crucial to their overall success. However, the selection of individuals is a potentially sensitive undertaking, as organizations try to select the best employees and also to maintain a diverse workforce. Organizations commonly go to great lengths to ensure that their selection instruments produce as little adverse impact against various protected groups as possible. Alleged violations of the Civil Rights Act (1991), Age Discrimination in Employment Act (1967) and the Americans with Disabilities Act (1990) are continually being tried in the courts. Therefore, organizations are concerned with techniques or methods to reduce adverse impact in order to obtain a diverse workforce and to avoid costly and damaging lawsuits.

Within the legal environment, an issue which has been raised several times in discrimination cases is whether test-wiseness plays a role in how various groups differ on their test performance (e.g., Bridgeport Guardians, Inc. v. Members of the Bridgeport Civil Service Commission, 1973; EEOC v. County
of Allegheny and Commonwealth of Pennsylvania, 1981; Firefighters Institute for Racial Equality v. City of St. Louis, 1976; Jones v. United States District Court for the Southern District of New York, 1975; Shield Club v. City of Cleveland, 1974; United States of America v. City of Chicago, 1976; United States of America v. H.K. Porter Company, Inc., 1968, Vulcan Pioneers v. New Jersey Department of Civil Service, 1987, etc.) . While the mention of this issue is often cursory and superficial, it is by no means inconsequential. Given the potential impact that test-wiseness could have on employment testing, it is surprising that this issue has not been empirically examined in the industrial/organizational psychology literature. While educational research has shed light on various issues involving test-wiseness (c.f., Scruggs & Lifson, 1985), no research has explored the impact of this construct in an employment context. Consequently, questions regarding whether the findings based on student samples generalize to an employment context remains an empirical question (Berkowitz & Donnerstein, 1982). Therefore the focus of the current research is to examine whether ethnic group differences in test-wiseness do in fact exist using an employment setting. In addition, this research seeks to examine whether test-wiseness can be effectively trained. If in fact ethnic group differences do exist, this research seeks to explore whether training can be implemented to eliminate or reduce ethnic group differences in test-wiseness.

**Background**

The issue of test-wiseness has long been a subject of interest to educational researchers. In one of the first discussions of the issue, Millman,
Bishop, and Ebel (1965) defined test-wiseness as an individual’s “capacity to utilize the characteristics and formats of the test and/or the test taking situation to receive a high score. Test-wiseness is logically independent of the examinee’s knowledge of the subject matter for which the items are supposedly measured” (p. 107). Since then, various education researchers have examined the impact that test-wiseness has on students’ test performance. Of particular concern is previous research which has found evidence that test-wiseness may be a source of additional variance in test scores and a factor which may lower test validity (Dolly & Vick, 1986; Fagley, 1987). Given the impact that lower validity as a result of test-wiseness could have on employment testing, it is surprising that attention has not been focused on this construct within the field of industrial-organizational psychology. Therefore, the present research intends to examine test-wiseness in an employment context.

Test-wiseness is thought to be comprised of various aspects including time management skills, answering all questions, and making sure that one fully understands the directions of tests (Millman et al., 1965). In addition, characteristics of the items themselves are also related to test-wiseness. There are various cues that can be embedded within items, which lead individuals to correctly guess the answer without having actual knowledge about the item content. Therefore, items can include cues such as the use of information from other items, alliterative associations, longer correct alternatives, more specific correct alternatives, etc. This paper will focus directly on embedded cues that
may help individuals decipher the correct alternative without relying on knowledge or experience with the content area.

Relatively little research has looked at measurement issues related to test-wiseness. An exception to this, however, is research conducted by Miller, Fuqua, and Fagley (1990). Using the Gibb Experimental Test of Test-Wiseness, a widely used measure of test-wiseness in the educational field, the authors conducted a factor analysis. This analysis revealed that there appear to be two dimensions underlying the concept of test-wiseness. While the authors did not label these factors, the types of items which load on each factor are as follows. Items which contain cues such as alliterative associations, more precise alternatives, longer correct alternatives and grammatical cues loaded on factor one. Items containing cues such as grossly unrelated alternatives, inclusionary language, and give-aways in other items loaded on factor two. Recently, Harmon, Morse, and Morse (1996) performed a confirmatory factor analysis on the Gibb Experimental Test of Test-Wiseness in order to assess the stability of a two-factor model of test-wiseness. Their results indicated support for a two-factor model.

In the test-wiseness literature, there are several variables that are thought to influence or be highly correlated with test-wiseness. One of these variables is test-taking experience (Sarnacki, 1979). For example, Kreit (1968) found evidence that third graders’ intelligence test scores improved after taking three different intelligence tests. This research implies that relatively little experience is necessary to learn test taking skills. However, as Sarnacki (1979) points out,
such experience is tempered by a number of factors...Mere experience in testing does not guarantee future success on tests, nor does it qualify an examinee as a skilled test-taker.” (p. 264). However, it is also thought that the amount of time that has elapsed since individuals had test taking experience influences test-wiseness (Sarnacki, 1979). In an overview of test-wiseness, Sarnacki (1979) describes research by Woodley (1973) and Bajtelsmit (1975) which found that adult subjects appeared to lacking in test taking skills and conjectured that this lack of skills was due to the lack of recent exposure to tests. However, researchers have found that test-wiseness skills span across a wide range of ages, including pre-school children (Gaines & Jongsma, 1974; Oakland, 1972), grade school children (Callenbach, 1973; Diamond & Evans, 1972; Kreit, 1968), junior and senior high school students (Crehan et al., 1974; Gross, 1976; Slakter, Koehler, & Hampton, 1970), and college students (Pryczak, 1973). In addition, Bajtelsmit (1975) and Woodley (1973) were able to teach adults how to effectively use test-wiseness strategies on multiple choice tests.

Within the literature of test-wiseness, a controversy exists concerning the degree to which test-wiseness correlates with general cognitive ability. Several studies have found support for the notion that test-wiseness and cognitive ability are separate issues. For example, Miguel (1997) found that even when the effects of general mental ability were controlled for, test-wiseness was still a significant predictor of performance on reading comprehension questions without the passages. Similarly, Sarnacki (1979) describes an unpublished masters thesis by Ardiff (1965) which found low positive correlations between test-
wiseness and cognitive ability in studies using third and sixth grade students and Dunn and Goldstein (1959) found zero correlations between cognitive ability and test-wiseness. These studies lend support to the idea that test-wiseness and general mental ability are two separate constructs. In addition, Crehan, Gross, Koehler, and Slakter (1978) reported that test-wiseness is not highly related to cognitive ability. Others have reported that test-wise individuals often score higher than those low in test-wiseness who are equal in terms of cognitive ability (Gross, 1977; Wahlstrom & Boersma, 1968). Therefore, many have argued that test-wiseness and cognitive ability are “moderately correlated at best” (Sarnacki, 1979).

Scruggs and Lifson (1985), however, argue that test-wiseness and cognitive ability are more closely related than others have argued. For example, Scruggs and Lifson contend that there is a lack of substantial evidence to support that cognitive ability and test-wiseness are separate constructs, and state that it “would defy credibility to assert that these ‘deductive reasoning’ strategies are not related to general mental ability” (p.342). In their argument, Scruggs and Lifson (1985) cite findings by Anderson (1973) which found a significant yet moderate correlation between test-wiseness and general mental ability, as did Diamond and Evans (1972). Based on these findings, Scruggs and Lifson (1985) claim that test-wiseness is not a construct that “students happen to acquire by chance or serendipity, which is unrelated to intelligence, and which results in substantial fluctuations of scores in achievement tests” (p. 342).
Is Test-wiseness Research Generalizable to Employment Settings?

There has been a long history of debate within psychology regarding the use of students as subjects in scientific investigations (Berkowitz & Donnerstein, 1982; Dobbins, Lane, & Steiner, 1988a; Dobbins, Lane, & Steiner, 1988b; Gordon, Slade, & Schmitt, 1986; Slade & Gordon, 1988). In particular, critics of laboratory research have questioned whether findings based on undergraduate students can be generalized to employment settings (Dobbins, et al., 1988).

Within the educational field, test-wiseness has been examined in terms of its impact on student performance. From an educational perspective, the use of student samples makes perfect sense. However, it should not be assumed that the findings from this domain would generalize to a population of job applicants or employees. Whether laboratory results generalize to other situations is an empirical question that should be addressed (Berkowitz & Donnerstein, 1982; Gordon, Slade & Schmitt, 1986).

One explanation of why students and employees may differ involves potential motivational differences. Gordon, Slade and Schmitt (1986) argue that there are a number of background variables that may influence how participants perceive the experimental task. Therefore, various motivational differences may exist among individuals. For example, Latham and Dossett (1978) describe how different perceptions of importance may exist for students who engage in temporary part-time work and workers who rely on their jobs for a living. Also, Jennings (1953) found that individuals taking a promotional test scored higher than individuals who take the same test for research purposes. Moreover,
significant motivational differences between job incumbents and applicants have been found (Arvey, Strickland, Drauden, & Martin, 1990). Therefore, it is possible that students would differ from actual job applicants in their motivation to learn test-wiseness skills in a training program.

Given the potential differences between students and applicant populations, the question still remains whether research involving test-wiseness which has been conducted in educational settings can be generalized. The present research seeks to examine this question by looking at whether or not the research on test-wiseness involving student samples is, in fact, generalizable to a sample of job applicants and employees. Consequently, the following discussion will investigate several issues which have been debated for quite some time in the educational literature. In addition, the implications of these findings will be discussed from an industrial-organizational psychology perspective. It is hoped that research directly exploring these issues in an employment setting will provide a greater understanding of such implications.

Can Test-wiseness Be Effectively Trained in an Employment Setting?

Within educational research, evidence exists which indicates that individuals high in test-wiseness have a greater chance of correctly responding to a “test-wise susceptible item” than those lower in test-wiseness (Rogers & Bateson, 1991). Therefore, a necessary question involves whether this skill can be trained. Previous research involving test-wiseness training has found evidence that participants can learn test-taking strategies that help them make more accurate guesses (Dolly & Williams, 1985). For example, Dolly and Vick
(1986) trained participants on test-wiseness and found that this training improved their performance on subsequent tests. Various other researchers have found evidence of training effects on the Metropolitan Readiness Test (Oakland, 1972), the Stanford Reading Test (Callenbach, 1973) and the Iowa Tests of Educational Development (Omvig, 1971). However, the focus of the test-wiseness skills that were trained differed across these studies with many of them focusing on issues such as using time effectively, error avoidance and reasoning strategies (Sarnacki, 1979). Interestingly, Langer, Wark, and Johnson (1973) found evidence that any type of instruction on item cues resulted in higher test-wiseness scores than no training at all.

Within this research, however, others have argued that the findings have been inconclusive regarding whether test-wiseness is a skill that can be effectively trained. In particular, Scruggs and Lifson (1985) argue that claims which suggest that test-wiseness strategies can be taught in a relatively short amount of time and can result in significantly higher performance were inflated and they argue that what is important is not statistically significant changes, but relative effect size. In particular, Scruggs and Lifson cite a meta-analysis by Bangert-Downs, Kulik, & Kulik (1983) which exclusively incorporated studies on student populations. The findings from this meta-analysis of educational studies indicated that the average effect size on achievement test scores following test-wise training was .29, which would translate into approximately three months of academic achievement. Scruggs and Lifson regard this as not a large difference and cite an additional meta-analysis (Scruggs, White, & Bennion, 1986) that
found that the average effect size for raising scores on achievement tests through test-wiseness training was .10. In addition, Samson (1985) conducted a meta-analysis and concluded that training programs which continue for five or more weeks result in significantly greater improvements than shorter programs. Again, however, the research on which this meta-analysis was based included only studies involving students. Therefore, the question still remains regarding how job applicants may differ in terms of their ability to be trained as well as what implications any changes in test-wiseness may have in a selection situation.

**Group Differences and Test-wiseness**

Given that selection tests are continually plagued with issues related to adverse impact, strategies which may prove to reduce adverse impact are continually being examined (Barrett, Doverspike, Cellar, & Johnson, 1991). Therefore, if ethnic group differences exist in test-wiseness, and test-wiseness is a skill that can be trained, this raises a potential ethical and legal issue (Miguel, 1997). According to the American Psychological Association’s Standards for Psychological Testing (1985), test-taking strategies which are unrelated to test content should be explained to individuals before the test is given, especially if these strategies have been found to significantly impact test performance. However, as Miguel (1997) notes, this standard is rarely followed in most employment testing situations.

Given the research reviewed above which has reported the beneficial effects of test-wiseness training, it is quite possible that individuals who are lower in test-wiseness may be at a relative disadvantage in a testing environment
(Crehen, Koehler, & Slatker, 1974; Moore, Schultz, & Baker, 1966). In particular, if there are ethnic group differences in test-wiseness such that African Americans tend to score lower in test-wiseness than whites, African Americans will be at a disadvantage in a testing environment. In other words, ethnic group differences in test-wiseness could mean that individuals in certain groups would be better able to answer questions without relying on prior knowledge or experience. This could have potentially important consequences in the area of employment selection and promotional exams.

The literature involving test-wiseness has debated the issue of whether ethnic group differences do in fact exist (Scruggs & Lifson, 1985). Several researchers have reported evidence of ethnic group differences. For example, in examining students’ abilities to answer reading comprehension items without the passages, Miguel (1997) and Barrett, Miguel, and Doverspike (in press) found evidence of black/white differences. Several other researchers have also reported ethnic group differences in test-wiseness. For example, Diamond, Ayres, Fishman, and Green (1976) and Ebel (1968) reported test-wise differences between African Americans and whites, Dreisbach and Keogh (1982) reported differences between whites and Hispanics, and Kalechstein et al., (1988) reported differences between whites and Asians.

Scruggs and Lifson (1985), however, argue that evidence which has been used to support the idea that minorities are lacking in test-wiseness is often unconvincing. In their discussion of a study by Diamond, Ayres, Fishman and Green (1976) which directly investigated the question of whether minority
populations differ in terms of test-wiseness, Scruggs and Lifson (1985) state that the results of the study do not directly demonstrate that disadvantaged or minority groups are lower in test-wiseness nor that a relationship exists between test-wiseness and achievement test scores for disadvantaged or minority groups. In addition, a review by McPhail (1984) concluded that research studies conducted on minority student populations have been inconclusive. Therefore, despite the concern raised by many researchers (e.g., Ebel, 1968) that score differentials may be due to group deficits in test-wiseness, relatively little research has focused on identifying this deficit.

In addition, much of the research exploring ethnic group differences in test-wiseness can be criticized on methodological grounds. For example, Kalechstein et al. (1988) discussed previous research that described the lack of test-wiseness in culturally different/disadvantaged groups. However, they administered a training program only to a group of African American disadvantaged second graders without comparing performance to a supposedly “advantaged” or white group. Therefore, their findings that these students were lacking test-wiseness skills may be due to the fact that all second graders are relatively inexperienced with tests and may all be lacking in test-wiseness skills. Consequently, it is not possible to infer that how African American second graders perform in the absence of a comparison group provides information regarding ethnic group differences in test-wiseness (Scruggs & Lifson, 1985). In addition, other studies examining the relationship of ethnic group and test-
wiseness have failed to come up with significant findings (e.g., Benson, Urman, & Hocevar, 1986; Diamond, Ayres, Fishman, & Green, 1976; Yearby, 1975).

From the above discussion, it is apparent that the issue of whether there are ethnic group differences in test-wiseness is still a debatable issue, and as Benson, Urman, and Hocevar (1986) pointed out, there is a relative lack of research that specifically focuses on minority groups and test-wiseness. In addition, whether such differences exist in an employment context is still a question that needs to be addressed. In particular, if such ethnic group differences exist, this may pose a definite disadvantage for minority applicants. A final issue involves the question that if ethnic group differences do in fact exist, is training able to effectively reduce or eliminate these differences?

Overview and Hypotheses

In summary, the present research investigates whether ethnic group differences exist in test-wiseness, whether test-wiseness is a skill that can be effectively trained, and whether training can help to reduce or eliminate any ethnic group differences. Three of the criteria from Kirkpatrick’s (1959) taxonomy are used to evaluate the effectiveness of test-wiseness training in a selection environment. Reactions are assessed by asking participants whether they enjoyed the training and whether they felt it was enjoyable and effective. Learning is assessed by looking at changes in scores between a pre-test and a post-test following the training session. Behavior is determined by changes in scores between the pre-test and post-test on a measure containing items with
embedded test-wise cues. Therefore, the research proposed the following hypotheses:

**Hypothesis 1:** Test-wiseness training will be related to significant improvements in participants' performance.

a) Test-wise training will have a positive effect on participants’ reactions, which would be indicated through positive ratings following the training program.

b) Test-wise training will have a significant effect on participants’ ability to identify the strategies learned in training. This will be determined by directly asking subjects their knowledge of test-wise strategies using a 7-item multiple-choice measure. Subjects’ performance on this measure prior to training as well as after training will be assessed to determine whether any improvements occur.

c) Test-wise training will have a significant effect on the ability of participants to identify test-wise cues on a test of items that could not be answered based on prior knowledge or experience. These items will have test-wise cues embedded within them. Subjects’ performance on this measure prior to training as well as after training will be assessed to determine whether any improvements occur.

**Hypothesis 2:** A significant difference between African Americans and whites will exist on pre-test scores of test-wiseness.

a) African Americans will score significantly lower on the learning pre-test of test-wiseness (direct measure of knowledge of test-wiseness strategies).

b) African Americans will score significantly lower on the behavior pre-test of test-wiseness (items with test-wiseness cues embedded within them).

c) If ethnic group differences exist, test-wiseness training will significantly reduce this difference (see Figure 1).
Overall, the major focus of this research is to examine the efficacy of test-wiseness training in an employment context. The relationship of ethnic group and test-wiseness will also be explored. In addition, this research seeks to explore whether ethnic group differences can be eliminated or reduced through test-wiseness training.
CHAPTER III

METHOD

Subjects

122 firefighters from three different metropolitan areas were utilized in the present study. Subjects were obtained through a voluntary sign-up sheet to participate in a training program on assessment centers for selection and promotional purposes. Thirteen of these subjects were eliminated because they did not complete the biographical information sheet. Of the 109 remaining subjects, an additional 20 were eliminated from analyses because they failed to complete at least half of the items on each of the learning or behavior measures. Therefore, when respondents did not complete at least four items on both the pre and post learning measures, and at least nine items on both the pre and post behavior measures, they were not included in subsequent analyses. For the purposes of this research, only the whites and African Americans were included in the analyses, which resulted in the elimination of one Asian and one Hispanic. This resulted in a total of 87 subjects (65 whites and 22 African Americans). There were 3 women (3%) and 84 men (97%). Additional demographic information is presented in Table 1.
Table 1

*Means and standard deviations for demographics variables*

<table>
<thead>
<tr>
<th></th>
<th>Overall Sample (n=87)</th>
<th>African Americans (n=22)</th>
<th>Whites (n=65)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>t</td>
</tr>
<tr>
<td>Age</td>
<td>38.88</td>
<td>10.52</td>
<td>1.85</td>
</tr>
<tr>
<td>Work Experience</td>
<td>14.54</td>
<td>8.56</td>
<td>.87</td>
</tr>
<tr>
<td>Education (in years)</td>
<td>13.42</td>
<td>1.51</td>
<td>2.71 **</td>
</tr>
</tbody>
</table>

*Note:* The t value refers to a test of statistical significance between whites and African Americans. 
*p < .05.  **p < .01.
Procedure

As part of a larger eight hour training program on assessment centers, subjects were given two measures to assess their level of test-wiseness (learning and behavior pre-tests). Subjects were then instructed on test-wise strategies involving item construction. Following this training, subjects were given a measure to assess their reactions to the training program as well as two post-test measures (learning and behavior). Figure 2 outlines the measures used and the design of the study.

Figure 2

Overview of measures used in method

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biographic Information Sheet</td>
<td>7-Item Learning Pre-test Measure</td>
<td>Training Intervention</td>
<td>Reactions</td>
</tr>
<tr>
<td>17-Item Behavior Pre-test Measure</td>
<td>7-Item Learning Post-test Measure</td>
<td>17-Item Behavior Post-test Measure</td>
<td>Job Knowledge Test</td>
</tr>
</tbody>
</table>

During the overall training program, subjects were informed of the correct procedure for filling out a biographical information sheet. Subjects were instructed to fill in the information as if they were in an actual testing situation.
The biographical information sheet included information regarding race, sex, education, and years experience in civil service work. These measures can be found in Appendix A. In addition, three weeks after the training session, subjects completed a job knowledge test which was an actual test used for employment.

Measures

**Pre-test behavior measure.** A pre-test was given to assess participants’ knowledge of test-wise strategies prior to the training session. This pre-test contained two components: a Learning component and a Behavior component. During the pre-test the Behavior component was given first so that information from the Learning component did not contaminate the subjects’ responses. The Behavior component contained 21 items with three alternatives. These items were related in content to fire fighter positions, yet were fictional in nature. In other words, there were no correct answers to these items. Therefore, subjects were not able to rely on past knowledge or experience in order to answer the items. Subjects were informed that the items were fictional and that there were no correct answers. Subjects were told to guess which alternative they felt was correct based on a test-wiseness cue. Each item contained one test-wise cue, and there were three questions for each of the seven cues identified by Gibb (1964). These cues are: grammatical cues, alliterative associations, longer correct alternatives, more precise correct alternatives, grossly unrelated alternatives, inclusionary ( absolutes) language, and give-aways in other items (see Table 2). Those participants who were knowledgeable in test-wise
strategies were expected to rely on these cues or test-wise strategies, while the remaining participants were expected to rely on idiosyncratic guessing methods.
Table 2

Overview of cues used in behavior measures

<table>
<thead>
<tr>
<th>Cue</th>
<th>Description</th>
<th>Sample Item</th>
</tr>
</thead>
</table>
| Grammatical             | Some items may contain grammatical errors or inconsistencies which can help indicate the correct alternative. These include errors involving subject verb agreement, use of plurals and singulars, etc. | Firefighter Jones has just finished his monthly review of how to properly wear oxygen tanks. Firefighter Jones learned that in order to safely ensure that one gets the correct supply of oxygen through his mask, **he should**:  
A. *screw* a TSR into the tank.  
B. *hooks* up the TSR gauge.  
C. *assembled* the TSR meter. |
| Alliterative Association| Identify a correct alternative because it sounds similar to a word in the stem of the question. | Firefighter Jones should treat a victim with a **Mellite** burn with:  
A. Dalfrexis.  
B. Bulofoid.  
C. Melproxin. |
| Longer Alternative      | Alternatives which are longer are often correct because the item writer wanted to make sure that all relevant or important information was included. | Upon arriving at the scene, Firefighter Jones pulls the fire engine to where the injured fire victims are being treated by the paramedics. Firefighter Jones knows that he should:  
A. park near the victims.  
B. navigate around the victims.  
C. **maneuver the engine between the fire and the victims**. |
### Table 2 (Continued)

| More Precise Alternative | Alternatives which contain more detail or are more precise are often correct because the item writer wanted to make sure that all relevant or important information was included. | When using Halon to fight a category 8 fire, Firefighter Jones should first ensure that:  
B. the hydraulic pressure is adequate.  
C. **the stream includes 20% cryptine.**  
D. fire personnel have proper safety equipment. |
|---------------------------|-------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|
| Grossly Unrelated Alternatives | The correct alternative to an item is determined by eliminating other alternatives. Specifically, some alternatives may be grossly unrelated to the topic of the item. These alternatives can then be eliminated which improves chances of guessing correctly. | Firefighter Jones was at a conference on combat strategies for firefighters. During the conference, Firefighter Jones learned that the city with the longest average response time to a fire in 1975 was:  
A. California.  
B. New Mexico.  
C. **Dallas.** |
| Inclusionary Language (Absolutes) | Involves avoiding certain key words, or absolutes within alternatives. Such words often imply that an alternative is incorrect because these words are very broad and difficult to defend. Such words include always, never, all, none, everyone, and nobody. | When attending a training session on the treatment of burn victims, Firefighter Jones learns that:  
A. burn victims respond well to desopin.  
B. **all** burn victims require nexolin.  
C. cryolin should **never** be given to burn victims. |
Table 2 (Continued)

<table>
<thead>
<tr>
<th>Give-Aways</th>
<th>Sometimes you may find clues or information in other questions within the test that may help you answer a particular question. By carefully reading each item, you may discover that some items contain similar information. In these situations, you may be able to find the answer to one item in a different item in the test.</th>
<th>During a training seminar on paramedic procedures, Firefighter Jones examined a slide of polydesmorphic neulukocyn. He learned that this substance is found in:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>When Firefighter Jones is examining a trauma patient, he should be aware that the normal percentage of polydesmorphic neulukocyn found in the <strong>blood</strong> of a healthy human is:</td>
<td>When Firefighter Jones is examining a trauma patient, he should be aware that the normal percentage of polydesmorphic neulukocyn found in the <strong>blood</strong> of a healthy human is:</td>
</tr>
<tr>
<td></td>
<td>A. 53/260 B. 2% C. 115%</td>
<td>A. 53/260 B. 2% C. 115%</td>
</tr>
</tbody>
</table>
Pre-Test learning measure. The Learning component was administered following the Behavior component. The Learning component was composed of seven items designed to directly assess subjects’ understanding of test-wise cues. For example, “Which of the following provides a clue that an alternative may be correct?”. The learning measure items may be found in Appendix B.

Reactions. Participants’ reactions were assessed through items contained on the post-test measure. Participants were asked the following two items: “How much did you enjoy the training program?” and “How effective do you feel the training program was?”. These were rated on a five-point Likert scale, ranging from “not at all” to “extremely”. These items were averaged to obtain an overall reaction score. The internal consistency of this measure was .85.

Post-test learning measure. The post-test was identical in form to the pre-test measure of test-wiseness and also included the Learning and Behavior components. During the post-test the Learning measure was administered first. The Learning component indicated the level of learning that occurred during the training program and included the same seven-items in the pre-test which directly assessed the subjects’ knowledge of test-wiseness strategies.

Post-test behavior measure. The second component, the Behavior measure included 21 items which involve issues related to fire fighting procedures but were fictional in nature. The actual content of the items on the post-test was different from the pre-test, but the items themselves contained the same test-wise cues that were in the pre-test items.
Biographical Information Sheet. The biographical information sheet was a computerized scan sheet containing items related to demographic information. Using a pencil, participants darkened circles that corresponded to information which most closely matched their demographic characteristics. Items included information regarding participants’ sex, race, age, education, and years of work experience.

Job Knowledge Test. The Job Knowledge Test consisted of 100 three-alternative multiple-choice items. Prior to the test, applicants were given a reading list of materials that were covered on the Job Knowledge Test. Items were written based on the information contained in the sources on the reading list. Applicants were administered this test in groups by a trained test proctor three weeks after the training session and was used to make actual hiring decisions.

Scale Development

The 7-item learning measure was created by the author in order to assess whether participants had learned the seven test-wiseness cues. These items are multiple choice items with three alternatives. Alternatives were chosen so that they were plausible and did not violate any test-wiseness cues. The behavior measures were created by a pool of professional item writers familiar with creating tests for firefighters. Item writers were instructed to create multiple choice items with three alternatives. In addition, they were instructed to create items based on fictional information so that no alternative was factually correct. Finally, they were instructed to embed one of the seven test-wiseness cues into
the item. In order to ensure that the pre-test and post-test items were equivalent, the items were pre-tested on a student sample.

A total of 109 subjects from three different colleges in the midwest participated in a classroom activity which entailed completing all 21 items from the pre-test and all 21 items from the post-test with no training intervention. To assess whether there were any order effects, 52 subjects completed the pre-test items first followed immediately by the post-test, and 57 subjects received the post-test items first followed immediately by the pre-test. Following the administration, the materials were collected. The subjects were then debriefed and given a demonstration of the test-wiseness training. All subjects were naïve to the profession of fire fighting.

This pilot test was performed to evaluate the properties of the items prior to the actual study with fire fighters. The decision rules that were used included 1) any items with a p value greater than .85 would be dropped, and 2) any items with a p value less than .15 would be dropped. These decision rules ensured that if 85% of naïve subjects got an item correct or 85% got the item wrong, the item was eliminated. The rationale of these decision rules was that if such a high percentage of naïve subjects were to get the item right, there was a greater possibility that an additional clue existed to make the item obvious to the subjects. Conversely, if a high percentage got the item wrong, there was a greater possibility that a different cue that was leading subjects to respond to a different alternative. Using these decision rules, one item from the pre-test was eliminated because it had a p-value greater than .85.
Tables 3 and 4 show the p values for each of the items on the pre-test and post-test behavior measures. Concurrently, two subject matter experts (members of a neighboring fire department that were not included in the study) were asked to evaluate the tests to ensure that they were truly fictional and that there were in fact no items that a firefighter would know based on experience. Based on their evaluations, a total of two items were dropped from the pre-test and two items were dropped from the post test.
Table 3

Pilot study p values for pre-test items counterbalanced for order effects

<table>
<thead>
<tr>
<th>Item Number</th>
<th>Overall P Value</th>
<th>P Value for Order One</th>
<th>P Value for Order Two</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (unrelated)**</td>
<td>.89</td>
<td>.94</td>
<td>.84</td>
</tr>
<tr>
<td>2 (grammatical)</td>
<td>.60</td>
<td>.50</td>
<td>.70</td>
</tr>
<tr>
<td>3 (alliterative)</td>
<td>.55</td>
<td>.56</td>
<td>.54</td>
</tr>
<tr>
<td>4 (most precise)*</td>
<td>.17</td>
<td>.17</td>
<td>.16</td>
</tr>
<tr>
<td>5 (longest)</td>
<td>.51</td>
<td>.50</td>
<td>.53</td>
</tr>
<tr>
<td>6 (give-away)</td>
<td>.75</td>
<td>.79</td>
<td>.70</td>
</tr>
<tr>
<td>7 (longest)</td>
<td>.23</td>
<td>.17</td>
<td>.28</td>
</tr>
<tr>
<td>8 (grammatical)</td>
<td>.49</td>
<td>.50</td>
<td>.47</td>
</tr>
<tr>
<td>9 (absolute)</td>
<td>.33</td>
<td>.29</td>
<td>.37</td>
</tr>
<tr>
<td>10 (give away)</td>
<td>.72</td>
<td>.81</td>
<td>.63</td>
</tr>
<tr>
<td>11 (absolute)</td>
<td>.80</td>
<td>.85</td>
<td>.75</td>
</tr>
<tr>
<td>12 (most precise)</td>
<td>.21</td>
<td>.21</td>
<td>.20</td>
</tr>
<tr>
<td>13 (alliterative)</td>
<td>.74</td>
<td>.75</td>
<td>.73</td>
</tr>
<tr>
<td>14 (absolute)</td>
<td>.47</td>
<td>.56</td>
<td>.38</td>
</tr>
<tr>
<td>15 (unrelated)</td>
<td>.83</td>
<td>.83</td>
<td>.84</td>
</tr>
<tr>
<td>16 (unrelated)</td>
<td>.54</td>
<td>.50</td>
<td>.58</td>
</tr>
<tr>
<td>17 (longest)</td>
<td>.48</td>
<td>.42</td>
<td>.54</td>
</tr>
<tr>
<td>18 (give-away)*</td>
<td>.68</td>
<td>.65</td>
<td>.70</td>
</tr>
<tr>
<td>19 (grammatical)</td>
<td>.84</td>
<td>.87</td>
<td>.81</td>
</tr>
<tr>
<td>20 (most precise)</td>
<td>.32</td>
<td>.35</td>
<td>.30</td>
</tr>
<tr>
<td>21 (alliterative)***</td>
<td>.78</td>
<td>.83</td>
<td>.74</td>
</tr>
<tr>
<td>Overall (all items)</td>
<td><strong>.57</strong></td>
<td><strong>.57</strong></td>
<td><strong>.56</strong></td>
</tr>
<tr>
<td>Overall (without dropped items)</td>
<td><strong>.55</strong></td>
<td><strong>.56</strong></td>
<td><strong>.55</strong></td>
</tr>
</tbody>
</table>

Note: n=52 for Order One; n=57 for Order Two. Order One refers to the condition where individuals received the pre-test items prior to post-test items; Order Two refers to the condition where individuals received the post-test items first followed by the pre-test items.

* - Item dropped because of SME judgments
** - Item dropped because combined p value > .85
*** - Item dropped to equate number of items on the pre and post tests
<table>
<thead>
<tr>
<th>Item Number</th>
<th>Overall P Value</th>
<th>P Value for Order One</th>
<th>P Value for Order Two</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (grammatical)</td>
<td>.46</td>
<td>.46</td>
<td>.46</td>
</tr>
<tr>
<td>2 (give-away)*</td>
<td>.68</td>
<td>.58</td>
<td>.79</td>
</tr>
<tr>
<td>3 (most precise)***</td>
<td>.16</td>
<td>.15</td>
<td>.18</td>
</tr>
<tr>
<td>4 (alliterative)</td>
<td>.76</td>
<td>.77</td>
<td>.75</td>
</tr>
<tr>
<td>5 (give-away)</td>
<td>.42</td>
<td>.33</td>
<td>.51</td>
</tr>
<tr>
<td>6 (give-away)</td>
<td>.59</td>
<td>.54</td>
<td>.65</td>
</tr>
<tr>
<td>7 (most precise)</td>
<td>.27</td>
<td>.31</td>
<td>.23</td>
</tr>
<tr>
<td>8 (longest)</td>
<td>.28</td>
<td>.31</td>
<td>.25</td>
</tr>
<tr>
<td>9 (longest)</td>
<td>.70</td>
<td>.67</td>
<td>.72</td>
</tr>
<tr>
<td>10 (unrelated)</td>
<td>.50</td>
<td>.50</td>
<td>.49</td>
</tr>
<tr>
<td>11 (absolute)</td>
<td>.51</td>
<td>.44</td>
<td>.58</td>
</tr>
<tr>
<td>12 (most precise)</td>
<td>.18</td>
<td>.19</td>
<td>.16</td>
</tr>
<tr>
<td>13 (alliterative)</td>
<td>.52</td>
<td>.54</td>
<td>.51</td>
</tr>
<tr>
<td>14 (absolute)</td>
<td>.67</td>
<td>.69</td>
<td>.65</td>
</tr>
<tr>
<td>15 (absolute)</td>
<td>.31</td>
<td>.27</td>
<td>.35</td>
</tr>
<tr>
<td>16 (unrelated)***</td>
<td>.74</td>
<td>.71</td>
<td>.77</td>
</tr>
<tr>
<td>17 (longest)</td>
<td>.21</td>
<td>.19</td>
<td>.23</td>
</tr>
<tr>
<td>18 (grammatical)</td>
<td>.67</td>
<td>.58</td>
<td>.75</td>
</tr>
<tr>
<td>19 (alliterative)*</td>
<td>.32</td>
<td>.40</td>
<td>.23</td>
</tr>
<tr>
<td>20 (grammatical)</td>
<td>.68</td>
<td>.64</td>
<td>.73</td>
</tr>
<tr>
<td>21 (unrelated)</td>
<td>.81</td>
<td>.71</td>
<td>.91</td>
</tr>
<tr>
<td>Overall (all items)</td>
<td>.50</td>
<td>.48</td>
<td>.52</td>
</tr>
<tr>
<td>Overall (without dropped items)</td>
<td>.50</td>
<td>.48</td>
<td>.53</td>
</tr>
</tbody>
</table>

*Note:  n=52 for Order One; n= 57 for Order Two. Order One refers to the condition where individuals received the pre-test items prior to post-test items; Order Two refers to the condition where individuals received the post-test items first followed by the pre-test items.

* - Item dropped because of SME judgments

*** - Item dropped to equate number of items on the pre and post tests
Finally, in order to balance the pre and post test in terms of the number of items and the number of items per test-wiseness cue, one additional item from the pre-test, and two additional items from the post test were eliminated. This resulted in a total of 17 items on both the pre and post tests. Within each test, three items tapped each of the following cues: grammatical error, longest alternative, and the use of absolutes. Two items tapped the following cues: sounds similar (alliterative) alternative, more precise alternative, unrelated/im plausible alternatives, and give-aways from another item.

When the pre-test was administered first, the average p value was .56. When the post test was administered first, the average p value of the pre-test was .55. Therefore, the order effect appears to be negligible. When combined, the average p value was .55. Conversely, when the post-test was administered first, the average p value was .53; When administered second, the average p value was .48. The combined average p value was .50. Therefore, the post test appeared to be slightly more difficult than the pre-test based on the pilot sample. Table 5 provides an overview of the p values for each of the items included organized by cue, the overall p values for each cue and the overall p values for both the pre and post test measures.
Table 5

Pilot test p values for pre-test and post-test items by test-wiseness cue

<table>
<thead>
<tr>
<th>Grammar Cue</th>
<th>Sounds Similar</th>
<th>Longest Alternative</th>
<th>Most Precise Alternative</th>
<th>Unrelated Alternatives</th>
<th>Absolutes</th>
<th>Give-Away</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Test</td>
<td>(Average p value = .55)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.60 (Q.2)</td>
<td>0.55 (Q.3)</td>
<td>0.51 (Q.5)</td>
<td>0.21 (Q.12)</td>
<td>0.83 (Q.15)</td>
<td>0.33 (Q.9)</td>
<td>0.75 (Q.6)</td>
</tr>
<tr>
<td>0.49 (Q.8)</td>
<td>0.74 (Q.13)</td>
<td>0.23 (Q.7)</td>
<td>0.32 (Q.20)</td>
<td>0.54 (Q.16)</td>
<td>0.80 (Q.11)</td>
<td>0.72 (Q.10)</td>
</tr>
<tr>
<td>0.84 (Q.19)</td>
<td>0.48 (Q.17)</td>
<td></td>
<td></td>
<td></td>
<td>0.47 (Q.14)</td>
<td></td>
</tr>
<tr>
<td>Average p-value for cue</td>
<td>.64</td>
<td>.65</td>
<td>.41</td>
<td>.27</td>
<td>.69</td>
<td>.53</td>
</tr>
<tr>
<td>Post Test</td>
<td>(Average p value = .50)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.46 (Q.1)</td>
<td>0.76 (Q.4)</td>
<td>0.28 (Q.8)</td>
<td>0.27 (Q.7)</td>
<td>0.50 (Q.10)</td>
<td>0.51 (Q.11)</td>
<td>0.42 (Q.5)</td>
</tr>
<tr>
<td>0.67 (Q.18)</td>
<td>0.52 (Q.13)</td>
<td>0.70 (Q.9)</td>
<td>0.18 (Q.12)</td>
<td>0.81 (Q.21)</td>
<td>0.67 (Q.14)</td>
<td>0.59 (Q.6)</td>
</tr>
<tr>
<td>0.68 (Q.20)</td>
<td>0.21 (Q.17)</td>
<td></td>
<td></td>
<td></td>
<td>0.31 (Q.15)</td>
<td></td>
</tr>
<tr>
<td>Average p-value for cue</td>
<td>.60</td>
<td>.64</td>
<td>.40</td>
<td>.23</td>
<td>.66</td>
<td>.50</td>
</tr>
</tbody>
</table>

Note: n=109

Training Session

The training session was conducted as part of a larger training session on assessment center and testing procedures. The content of the larger training program involved issues such as listing the appropriate source materials for the job knowledge tests, giving example assessment center activities, and informing individuals of the place and time that they were supposed to report to the employment test.

The test-wiseness training was conducted within this larger training program by the author and one other individual professionally trained in selection procedures, test-wiseness, and item writing procedures. Individuals were trained...
in group settings ranging in size from 15 to 50 people. The training involved first
giving the participants the pre-test measures. Once these measures were
completed and collected, the participants were instructed as to what test-
wiseness was and what strategies they may use to try to effectively guess in
situations that they did not know the answer to a question.

This explanation included a series of handouts for the participants as well
as overhead slides, which explained each of the strategies (see Appendix D). In
addition, examples were provided to the participants to further explain the
strategies. Participants were encouraged to ask questions. Once the training
session was over, the participants were asked to put aside their materials and
complete the post-test measures. Participants were informed that participation
was completely voluntary and that the measures were confidential and would in
no way be used in the scoring of their actual employment tests. The total amount
of time for the training ranged from 45 minutes to an hour.
CHAPTER IV

RESULTS

The following results are organized into four main sections. First, reaction results are presented. This section includes descriptive statistics as well as correlations between reaction scores and other measures. Second, learning measure results are presented. On these measures, if respondents failed to provide an answer to an item, it was scored as incorrect. Within this section, descriptive statistics, overall results, and results by race are presented. In the third section, the behavior results are shown. These results also include descriptive statistics, overall results and results by race. Finally, the fourth section explores possible explanations for the findings obtained in the previous sections.

Reaction Results

Descriptive statistics for the reaction measure are reported in Table 6. Hypothesis 1a proposed that test-wiseness training would have a positive effect on participants’ reactions, such that they would report positive ratings following the training program. In general, this was supported; subjects were somewhat positive in their reactions to the training program with an average score of 3.30 out of a possible 5.00. When broken apart by race, whites had a mean reaction score of 3.29 and African Americans had a mean reaction score of 3.31. This
difference between whites and African Americans was not statistically significant (t=.08, p=.98).

Table 6

_Means and standard deviations for the reaction measure by race_

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>77</td>
<td>3.30</td>
<td>.71</td>
<td>1.00 – 4.50</td>
</tr>
<tr>
<td>Whites</td>
<td>56</td>
<td>3.29</td>
<td>.71</td>
<td>1.00 – 4.00</td>
</tr>
<tr>
<td>African Americans</td>
<td>21</td>
<td>3.31</td>
<td>.73</td>
<td>1.50 – 4.50</td>
</tr>
</tbody>
</table>

*Note:* Reaction item responses were made on a 5-point response scale.

Consistent with Alliger and Janek’s findings (1989), reactions did not significantly correlate with either the learning or behavior measures (see Table 7). In addition, reactions did not correlate with the demographic variables of age, education, or work experience.
Table 7
Descriptive statistics and intercorrelations among variables used in study

| Variable                        | N   | M    | SD  | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   |
|---------------------------------|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Reactions                       | 77  | 3.30 | .71 |     |     |     |     |     |     |     |     |     |
| Pre-test Learning Measure       | 87  | 5.97 | 1.16|     |     |     |     |     |     |     |     |     |
| Post-test Learning Measure      | 87  | 6.36 | .88 |     |     |     |     |     |     |     |     |     |
| Pre-test Behavior Measure       | 87  | 10.46| 2.55|     |     |     |     |     |     |     |     |     |
| Post-test Behavior Measure      | 87  | 10.89| 3.04|     |     |     |     |     |     |     |     |     |
| Age                             | 85  | 38.67| 10.50|     |     |     |     |     |     |     |     |     |
| Education                       | 86  | 13.41| 1.50|     |     |     |     |     |     |     |     |     |
| Work Experience                 | 84  | 14.44| 8.49|     |     |     |     |     |     |     |     |     |

*p < .05  **p < .01
Learning Results

Overall, the average score on the pre-test learning measure was 5.97 (SD=1.16) and the average score on the post-test learning measure was 6.36 (SD=.88). [See Table 7]. This improvement from the pre-test to the post-test indicates a statistically significant training effect (t=-3.43, p<.001), thus supporting hypothesis 1b. In order to calculate the effect size, the formula found in Dunlap, Cortina, Vaslow, and Burke (1996) was used in order to correct for the correlation between measures in a repeated measures design:

\[ d = t_c \sqrt{\frac{2(1-r)}{n}}^{1/2} \]

In this formula \( t_c \) refers to the t statistic for the correlated observations and \( r \) is the correlation across pairs of measures. Using this formula, the effect size was –.38 (see Table 8). These findings, therefore, support hypothesis 1b, which predicted that test-wiseness training would have a significant effect on participants' ability to identify the test-taking strategies learned in training. In addition, this effect size was higher than the average effect size of .29 in the meta-analysis by Bangert-Downs, Kulik, & Kulik (1983) or .10 found in the meta-analysis by Scruggs, White, and Bennion (1986).
Table 8

Repeated measures effect sizes for pre-test and post-test measures for overall sample and broken apart by ethnic group

<table>
<thead>
<tr>
<th></th>
<th>Overall Sample (n=87)</th>
<th>African Americans (n=22)</th>
<th>Whites (n=65)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t</td>
<td>d</td>
<td>t</td>
</tr>
<tr>
<td>Learning Measure</td>
<td>-3.43 **</td>
<td>-.38</td>
<td>-1.69</td>
</tr>
<tr>
<td>Behavior Measure</td>
<td>-1.30</td>
<td>-.15</td>
<td>.95</td>
</tr>
<tr>
<td>Grammatical cues</td>
<td>-3.43 **</td>
<td>-.44</td>
<td>.46</td>
</tr>
<tr>
<td>Sounds Similar</td>
<td>-5.51 **</td>
<td>-.83</td>
<td>-2.01</td>
</tr>
<tr>
<td>Longest Alternative</td>
<td>-1.02</td>
<td>-.15</td>
<td>.88</td>
</tr>
<tr>
<td>Precise Alternative</td>
<td>-3.17 **</td>
<td>-.44</td>
<td>-2.13 **</td>
</tr>
<tr>
<td>Unrelated Alternative</td>
<td>5.74 **</td>
<td>.71</td>
<td>3.18 **</td>
</tr>
<tr>
<td>Absolutes</td>
<td>.09</td>
<td>-.01</td>
<td>.18</td>
</tr>
<tr>
<td>Give-aways</td>
<td>4.08 **</td>
<td>.58</td>
<td>2.89 **</td>
</tr>
</tbody>
</table>

Note: t refers to the t-statistic using a paired samples t-test. d refers to the repeated measures effect size using the equation by Dunlap, Cortina, Vaslow, and Burke (1986)
*p < .05. **p < .01.

When broken apart by race (see Table 9), whites had a mean of 5.98 (SD=1.11) on the pre-test and 6.37 (SD=.94) on the post-test. African Americans had a mean of 5.91 (SD=1.34) on the pre-test and 6.36 (SD=.66) on the post-test (see Figure 3). There were no statistically significant differences between whites and African Americans on either the pre-test measure (t=-.26, p>.05) or the post-test measure (t=-.03, p>.05). To explore the interaction effects of ethnicity and training on the learning post-test performance, a repeated measures ANOVA was performed. The results revealed that there was not a significant interaction effect with ethnicity (F(1, 85) = .07, p>.05). Based on the above findings, hypothesis 2a was not supported in that African Americans did not score significantly lower on the learning pre-test of test-wiseness. In addition,
hypothesis 2c was not supported in that there was not a significant interaction
effect whereby training reduced ethnic group differences between African
Americans and whites. However, it should be noted that there was a potential
ceiling effect due to high means on both the pre and post learning measures.

Table 9

*Means and standard deviations of learning measures by race*

<table>
<thead>
<tr>
<th></th>
<th>Overall Sample (n=87)</th>
<th>African Americans (n=22)</th>
<th>Whites (n=65)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean  SD</td>
<td>Mean  SD</td>
<td>Mean  SD</td>
</tr>
<tr>
<td>Pre-test Learning</td>
<td>5.97  1.17</td>
<td>5.91  1.34</td>
<td>5.98  1.11</td>
</tr>
<tr>
<td>Post-test Learning</td>
<td>6.37  .88</td>
<td>6.36  .66</td>
<td>6.37  .94</td>
</tr>
</tbody>
</table>

Figure 3

*Results on pre- and post-tests for learning and behavior measures by ethnic groups*

Learning Measures

![Graph showing learning measures pre-test vs. post-test for African Americans and Whites]
Behavior Results

Overall, the mean score for the pre-test behavior measure was 10.46 ($SD=2.55$) and the mean for the post test was 10.89 ($SD=3.04$). [See Table 7]. While this showed a slight improvement overall on the post-test, this difference was not statistically significant using a paired samples t-test ($t=-1.30$, $p=.20$). In addition, the effect size for this change was -.15 (see Table 8). Therefore, while the findings were in the hypothesized direction, hypothesis 1c was not supported. This effect size was smaller than the .29 effect size found in the meta-analysis by Bangert-Downs, Kulik, & Kulik (1983), but larger than the .10 effect size found in the meta-analysis by Scruggs, White, and Bennion (1986). It is important to note, however, that the effect size is difficult to interpret given that the pre and post tests were not parallel forms.
When examined by race (see table 10), whites had a mean score of 10.49 (SD=2.76) and African Americans had a mean score of 10.36 (SD=1.84) on the pre-test behavior measure. On the post-test measures, whites had a mean score of 11.25 (SD=3.10) and African Americans had a mean score of 9.82 (SD=2.63). Therefore, while whites improved slightly, African Americans actually decreased slightly (see Figure 4). To explore the interaction effects of ethnicity and training on the behavior post-test performance, a repeated measures ANOVA was performed. The results revealed that the interaction was significant at a liberal alpha level of p<.10 (F(1, 85) = 3.06, p=.08).

Table 10

<table>
<thead>
<tr>
<th></th>
<th>Overall Sample (n=87)</th>
<th>African Americans (n=22)</th>
<th>Whites (n=65)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean  SD</td>
<td>Mean  SD</td>
<td>Mean  SD</td>
</tr>
<tr>
<td>Pre-test Behavior</td>
<td>10.46  2.55</td>
<td>10.36  1.84</td>
<td>10.49  2.76</td>
</tr>
<tr>
<td>Post-test Behavior</td>
<td>10.89  3.04</td>
<td>9.82  2.63</td>
<td>11.25  3.10</td>
</tr>
</tbody>
</table>

Similar to the findings of the learning measures, hypothesis 2b was not supported in that the differences between whites and African Americans was not statistically significant on the pre-test behavior measure (t=-.20, p>.05). However, the difference between whites and African Americans on the post test behavior measure was significant at a liberal alpha level of p<.10 (t=-1.94, p=.06). In contrast to hypothesis 2c, test-wiseness training did not appear to alleviate ethnic group differences between whites and African Americans. In fact
test-wiseness training appeared to slightly exacerbate group differences, with the interaction effect approaching significance.

**Behavior results by cue dimensions.** In addition to the overall score on the pre-test and post-test, subjects’ scores were evaluated in terms of the seven test-wiseness cues (see Table 11 for the breakdown of means by cues for whites and African Americans). Overall, there were statistically significant differences between the pre and post measures for grammatical cues ($t=-3.43$, $p<.001$), sounds similar cues ($t=-5.51$, $p<.001$), more precise alternative cues ($t=-3.17$, $p<.01$), unrelated alternatives cues ($t=5.74$, $p<.001$), and give away cues ($t=4.08$, $p<.001$). However, the change in scores on unrelated alternative cues and the give away cues was in the opposite direction than was expected. That is, subjects’ scores actually decreased on the post test. There were no statistically significant differences on the longest alternative cues ($t=-1.03$, $p>.05$), and absolutes ($t=.09$, $p>.05$). Table 12 shows the intercorrelations of each of the cue dimensions. In general, pre-post scores for the same cue effect did not correlate very highly. However, grammatical cues and unrelated alternatives did have significant correlations between the pre and post scores. The lack of strong pre-post cue effect correlations is likely due to the fact that there were few items in each scale (i.e., two or three items per scale) and, therefore, low reliability.
Table 11

Means and standard deviations of Cue Dimensions for behavior pre-test and post test by race

<table>
<thead>
<tr>
<th>Cue Dimension</th>
<th>Overall (n=87)</th>
<th>White (n=65)</th>
<th>African American (n=22)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Grammatical Cue</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
<td>.63</td>
<td>.27</td>
<td>.60</td>
</tr>
<tr>
<td>Post test</td>
<td>.75</td>
<td>.29</td>
<td>.77</td>
</tr>
<tr>
<td>Sounds Similar Cue</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
<td>.66</td>
<td>.29</td>
<td>.67</td>
</tr>
<tr>
<td>Post test</td>
<td>.89</td>
<td>.25</td>
<td>.90</td>
</tr>
<tr>
<td>Longest Alternative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
<td>.57</td>
<td>.30</td>
<td>.55</td>
</tr>
<tr>
<td>Post test</td>
<td>.61</td>
<td>.31</td>
<td>.64</td>
</tr>
<tr>
<td>Most Precise Alternative Cue</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
<td>.28</td>
<td>.35</td>
<td>.31</td>
</tr>
<tr>
<td>Post test</td>
<td>.44</td>
<td>.38</td>
<td>.44</td>
</tr>
<tr>
<td>Unrelated Alternative Cue</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
<td>.80</td>
<td>.29</td>
<td>.81</td>
</tr>
<tr>
<td>Post test</td>
<td>.59</td>
<td>.33</td>
<td>.59</td>
</tr>
<tr>
<td>Absolute Cue</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
<td>.58</td>
<td>.30</td>
<td>.60</td>
</tr>
<tr>
<td>Post test</td>
<td>.58</td>
<td>.31</td>
<td>.60</td>
</tr>
<tr>
<td>Give-away</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
<td>.82</td>
<td>.31</td>
<td>.84</td>
</tr>
<tr>
<td>Post test</td>
<td>.61</td>
<td>.39</td>
<td>.67</td>
</tr>
</tbody>
</table>

Note: Cue Dimensions scores were calculated by taking the mean p value for each dimension. This was done because of unequal number of items in different dimensions. Therefore, scores could range from 0 to 1.
<table>
<thead>
<tr>
<th>Variable</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>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Pre-test Grammatical Cue</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></td>
<td></td>
</tr>
<tr>
<td>2. Pre-test Sounds Similar Cue</td>
<td>.15</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></td>
</tr>
<tr>
<td>3. Pre-test Longest Alt. Cue</td>
<td>.24*</td>
<td>-.10</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>
<tr>
<td>4. Pre-test Precise Alt. Cue</td>
<td>.20</td>
<td>-.02</td>
<td>.22*</td>
<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>5. Pre-test Unrelated Alt. Cue</td>
<td>.13</td>
<td>.17</td>
<td>.06</td>
<td>.21*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Pre-test Absolute Cue</td>
<td>.17</td>
<td>-.02</td>
<td>.19</td>
<td>.06</td>
<td>.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Pre-test Give-away Cue</td>
<td>.19</td>
<td>.01</td>
<td>.09</td>
<td>.06</td>
<td>.08</td>
<td>.23*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Post-test Grammatical Cue</td>
<td>.30**</td>
<td>-.07</td>
<td>.14</td>
<td>.05</td>
<td>.17</td>
<td>.26**</td>
<td>.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Post-test Sounds Similar Cue</td>
<td>.05</td>
<td>.02</td>
<td>.03</td>
<td>-.09</td>
<td>.09</td>
<td>.21</td>
<td>.03</td>
<td>.32**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Post-test Longest Alt. Cue</td>
<td>.05</td>
<td>-.14</td>
<td>.05</td>
<td>.09</td>
<td>-.01</td>
<td>.22*</td>
<td>-.17</td>
<td>.02</td>
<td>.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Post-test Precise Alt. Cue</td>
<td>.19</td>
<td>.09</td>
<td>.12</td>
<td>.15</td>
<td>.14</td>
<td>.30**</td>
<td>.03</td>
<td>.19</td>
<td>.05</td>
<td>.26*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Post-test Unrelated Alt. Cue</td>
<td>.06</td>
<td>.01</td>
<td>.07</td>
<td>.04</td>
<td>.34**</td>
<td>.04</td>
<td>.07</td>
<td>.39**</td>
<td>.45**</td>
<td>.01</td>
<td>.21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Post-test Absolute Cue</td>
<td>.18</td>
<td>.03</td>
<td>.17</td>
<td>.31**</td>
<td>.28**</td>
<td>.18</td>
<td>-.07</td>
<td>.17</td>
<td>.09</td>
<td>.17</td>
<td>.22*</td>
<td>.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Post-test Give-away Cue</td>
<td>.21*</td>
<td>.02</td>
<td>.08</td>
<td>.25*</td>
<td>.15</td>
<td>.20</td>
<td>.13</td>
<td>.39**</td>
<td>.32**</td>
<td>.09</td>
<td>.25*</td>
<td>.13</td>
<td>.41**</td>
<td></td>
</tr>
</tbody>
</table>

Note: n=87

*p < .05. **p < .01.
When evaluating the white participants’ scores using paired samples t-tests for each of the cue dimensions, there were statistically significant differences between the pre-test and post test for grammatical cues ($t=-4.23, p<.001$), sounds similar (alliterative) cues ($t=-5.42, p<.001$), most precise cues ($t=-2.10, p<.05$), unrelated alternatives cues ($t=4.77, p<.001$), and give-away cues ($t=3.01, p<.01$). The difference between the pre-test and post-test was significant at a liberal alpha level of $p<.10$ for the longest alternative cues ($t=-1.94, p=.06$). As with the overall sample, subjects scored lower on the post-test for unrelated alternative cues and give away cues which is in the opposite direction than is expected. Only absolute cues did not see a statistically significant difference between the pre and post test measures.

Using paired samples t-tests for each of the cue dimensions, there were statistically significant differences between the pre-test and post test for African Americans on most precise alternative cues ($t=-2.13, p<.01$), unrelated alternatives, ($t=3.18, p<.01$) and give-aways ($t=2.89, p<.01$). The difference between pre-test and post test scores was significant at a liberal alpha level of $p<.10$ for sounds similar cues ($t=-2.01, p=.06$). Again, as with the overall sample and the white sample, African Americans scored lower on the post-test for unrelated alternative cues and give away cues, which is in the opposite direction than was expected. There were no statistically significant differences between pre and post test measures for African Americans on grammatical cues, longest alternatives, and absolutes.
Scores on the cue dimensions were evaluated to determine if there were any statistically significant differences between whites and African Americans. For all cues on the pre-test, there were no statistically significant differences between whites and African Americans, however, the difference was significant at a liberal alpha level of $p<.10$ on the grammatical cues ($t=1.70$, $p=.09$) with African Americans scoring higher than the whites. On the post-test, whites scored significantly higher than the African Americans on give away cues ($t=-2.31$, $p<.05$).

**Exploratory Analyses**

The effect of age. As can be seen in Table 7, age was significantly correlated with the post-test Behavior measure, but not with the pre-test. Similarly, while the correlation between age and the Learning post-test was not statistically significant, it was considerably larger than the correlation between age and the pre-test Learning measure. Given organizations’ need for concern regarding any adverse effects against individuals older than forty due to the Age Discrimination in Employment Act (ADEA), the data were separated into two groups, those forty and older ($n=48$) and those below forty ($n=39$). Table 13 shows the means for these groups on the Behavior and Learning measures. When looking at the Learning measures, subjects under forty had an effect size of -.17 due to training and subjects over forty had an effect size of -.56 (see Figure 4). This same pattern occurred with the Behavior measures, with an effect size of .17 for those under forty and an effect size of -.46 for those over forty (see Figure 4). In addition, there were no statistically significant differences
between those under forty and those over forty on either the Learning pre-test (t=-.12, p>.05) or the Behavior pre-test (t=-.42, p>.05), but there was a statistically significant difference between those under forty and those over forty on the Learning post-test (t=-2.36, p<.05) and the Behavior post-test (t=-3.09, p<.01).

Table 13

Means and standard deviations of learning and behavior measures by age group

<table>
<thead>
<tr>
<th></th>
<th>Overall Sample (n=87)</th>
<th>Under 40 (n=39)</th>
<th>Over 40 (n=48)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean  SD</td>
<td>Mean  SD</td>
<td>Mean  SD</td>
</tr>
<tr>
<td>Learning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
<td>5.97</td>
<td>1.17</td>
<td>5.95</td>
</tr>
<tr>
<td>Post-test</td>
<td>6.37</td>
<td>.88</td>
<td>6.13</td>
</tr>
<tr>
<td>Behavior</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
<td>10.46</td>
<td>2.55</td>
<td>10.33</td>
</tr>
<tr>
<td>Post-test</td>
<td>10.89</td>
<td>3.04</td>
<td>9.82</td>
</tr>
</tbody>
</table>

Figure 4

Results on pre- and post test learning and behavior measures by age groups

Learning Measure

![Graph showing results on pre- and post test learning and behavior measures by age groups]
Relationship with job knowledge test scores. For a subset of the sample (N=55), actual job knowledge test scores used to select or promote individuals were obtained. The average job knowledge test score was 73.95 (SD= 7.97) (see Table 14). When broken apart by race, African Americans had an average score of 70.53 (SD=6.75) and whites had an average score of 75.47 (SD=8.08). A t-test was conducted which indicated that the difference between African Americans and whites was statistically significant (t=-2.20, p<.05).

Table 14

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>55</td>
<td>73.95</td>
<td>7.97</td>
<td>56-92</td>
</tr>
<tr>
<td>Whites</td>
<td>38</td>
<td>75.47</td>
<td>8.08</td>
<td>57-92</td>
</tr>
<tr>
<td>African Americans</td>
<td>17</td>
<td>70.53</td>
<td>6.75</td>
<td>56-80</td>
</tr>
</tbody>
</table>
The intercorrelations of these test scores with demographic as well as
test-wiseness measures are shown in Table 15. Of the demographic variables,
the only variable that significantly correlated with job knowledge test scores was
age with younger individuals scoring higher. Interestingly, the correlations of job
knowledge test scores with the pre-test measures of test-wiseness (both learning
and behavior) are not statistically significant. However, there are statistically
significant correlations with both the post test learning measure ($r=.31$) and the
post test behavior measure ($r=.47$). When broken apart by race, only the whites
demonstrated a statistically significant correlation with the post test measures of
test-wiseness (see Table 16). However, there were only 17 African Americans
so power is definitely an issue.
Table 15: Descriptive statistics and intercorrelations of job knowledge test scores and test-wiseness and demographic variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>X</th>
<th>SD</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>
</tr>
</thead>
<tbody>
<tr>
<td>1. Job Knowledge Test</td>
<td>55</td>
<td>73.95</td>
<td>7.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Age</td>
<td>54</td>
<td>44.83</td>
<td>6.47</td>
<td>-0.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Education</td>
<td>55</td>
<td>13.60</td>
<td>1.57</td>
<td>0.06</td>
<td>0.25*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Work Experience</td>
<td>55</td>
<td>18.27</td>
<td>7.02</td>
<td>-0.16</td>
<td>-0.88**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Reactions</td>
<td>49</td>
<td>3.20</td>
<td>1.74</td>
<td>0.08</td>
<td>0.09</td>
<td>-0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Pre-test Learning Measure</td>
<td>55</td>
<td>5.98</td>
<td>1.25</td>
<td>0.19</td>
<td>0.19</td>
<td>0.01</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Post-test Learning Measure</td>
<td>55</td>
<td>5.62</td>
<td>1.65</td>
<td>0.31*</td>
<td>0.19</td>
<td>0.07</td>
<td>0.15</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Pre-test Behavior Measure</td>
<td>55</td>
<td>10.44</td>
<td>2.67</td>
<td>0.32**</td>
<td>0.47**</td>
<td>0.32**</td>
<td>0.32**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Post-test Behavior Measure</td>
<td>55</td>
<td>11.93</td>
<td>2.62</td>
<td>0.43**</td>
<td>0.41**</td>
<td>0.43**</td>
<td>0.43**</td>
<td>0.41**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05. **p < .01.
Table 16

Correlations between job knowledge test scores, test-wiseness variables, and demographic variables for whites and African Americans

<table>
<thead>
<tr>
<th>Job Knowledge Test Scores</th>
<th>Whites (n=38)</th>
<th>African Americans (n=17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reactions</td>
<td>-.26</td>
<td>.35</td>
</tr>
<tr>
<td>Pre-Learning</td>
<td>.23</td>
<td>.07</td>
</tr>
<tr>
<td>Post-Learning</td>
<td>.42*</td>
<td>-.15</td>
</tr>
<tr>
<td>Pre-Behavior</td>
<td>.24</td>
<td>-.22</td>
</tr>
<tr>
<td>Post-Behavior</td>
<td>.51*</td>
<td>.10</td>
</tr>
<tr>
<td>Age</td>
<td>-.21</td>
<td>-.35</td>
</tr>
<tr>
<td>Education</td>
<td>.26</td>
<td>-.39</td>
</tr>
<tr>
<td>Work Experience</td>
<td>-.21</td>
<td>.27</td>
</tr>
</tbody>
</table>

Job knowledge test scores were also correlated with each of the cues from the behavior measures (see Table 17). Using the overall sample that took the job knowledge test, scores were significantly correlated with the post-test measures for the longest alternative cues, most precise alternative cues, and the give-away cues. Again, when broken apart by race, these correlations were statistically significant only for the white individuals.

Given that age was significantly correlated with ethnic group (r=.24) with African Americans being older on average than whites and significant ethnic group differences on the job knowledge test, an ANCOVA model was tested that contained these two variables as covariates that may have a potential impact on the effects of training. While there was a significant effect of job knowledge test scores on learning post-test performance (F=3.90, p< .05), the expected interaction of ethnicity and improvement due to training on the Learning measure was not significant when controlling for the effects of age and job knowledge test scores (F (1,50) = .001, p> .05). However, there was a significant interaction
effect for ethnicity and improvement due to training on the behavior measure

\( F (1,50) = 4.21, p< .05 \) when controlling for the effects of age and job

knowledge test scores. Table 18 shows the means and standard deviations for

African Americans and whites on the behavior measures when controlling for the
effects of age and job knowledge test scores.

Table 17

Correlations between job knowledge test scores and test-wiseness cues for
whites and African Americans

<table>
<thead>
<tr>
<th>Job Knowledge Test Scores</th>
<th>Overall (N=55)</th>
<th>Whites (N=38)</th>
<th>African Americans (N=17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test Grammatical Cue</td>
<td>-.03</td>
<td>.12</td>
<td>-.42</td>
</tr>
<tr>
<td>Pre-test Sounds Similar Cue</td>
<td>.17</td>
<td>.14</td>
<td>.23</td>
</tr>
<tr>
<td>Pre-test Longest Alt. Cue</td>
<td>.03</td>
<td>.14</td>
<td>-.15</td>
</tr>
<tr>
<td>Pre-test Precise Alt. Cue</td>
<td>.16</td>
<td>.15</td>
<td>-.01</td>
</tr>
<tr>
<td>Pre-test Unrelated Alt. Cue</td>
<td>.21</td>
<td>.17</td>
<td>.22</td>
</tr>
<tr>
<td>Pre-test Absolute Cue</td>
<td>.09</td>
<td>.12</td>
<td>-.11</td>
</tr>
<tr>
<td>Pre-test Give-away Cue</td>
<td>.05</td>
<td>.14</td>
<td>-.28</td>
</tr>
<tr>
<td>Post-test Grammatical Cue</td>
<td>.10</td>
<td>-.04</td>
<td>.01</td>
</tr>
<tr>
<td>Post-test Sounds Similar Cue</td>
<td>.26</td>
<td>.28</td>
<td>.25</td>
</tr>
<tr>
<td>Post-test Longest Alt. Cue</td>
<td>.35 **</td>
<td>.38 *</td>
<td>.16</td>
</tr>
<tr>
<td>Post-test Precise Alt. Cue</td>
<td>.34 *</td>
<td>.37 *</td>
<td>.18</td>
</tr>
<tr>
<td>Post-test Unrelated Alt. Cue</td>
<td>.11</td>
<td>-.02</td>
<td>.15</td>
</tr>
<tr>
<td>Post-test Absolute Cue</td>
<td>.14</td>
<td>.27</td>
<td>-.28</td>
</tr>
<tr>
<td>Post-test Give-away Cue</td>
<td>.40 **</td>
<td>.44 **</td>
<td>.09</td>
</tr>
</tbody>
</table>

Table 18

Means and standard deviations for behavior measures by race controlled for age
and job knowledge test scores

<table>
<thead>
<tr>
<th></th>
<th>Overall Sample (n=87)</th>
<th>African Americans (n=22)</th>
<th>Whites (n=65)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Pre-test Behavior</td>
<td>11.98</td>
<td>2.61</td>
<td>10.31</td>
</tr>
<tr>
<td>Post-test Behavior</td>
<td>10.43</td>
<td>2.69</td>
<td>10.25</td>
</tr>
</tbody>
</table>
Student Data as a Non-Equivalent Control Group.

Due to the lack of a control group, whether differences in pre- and post-test scores is due to training or some other factor is difficult to determine. Therefore, a quasi-experimental approach was taken in which a non-equivalent control group was used. Goldstein (1986) recommends the use of a non-equivalent control group in situations where it is not possible to use a control group. In this study, the student sample acts as a non-equivalent control group because students were not given the training until after they had completed both the Behavior pre and post tests. When using a non-equivalent control group, the more similar the groups are on their pre-test scores, the greater the evidence of training effects if there are differences on the post test measures.

Table 19 shows the means for employees and students on the Behavior measures (students did not receive the Learning measures). Paired t-tests were performed to investigate whether these two groups differed. There was a statistically significant difference between Behavior pre and post-tests for students (t=2.94, p<.01) with students performing lower on the post-test. However, there was not a statistically significant difference between the pre and post tests for employees (t=-1.31, p>.05). In addition, there were statistically significant differences between students and employees on both the Behavior pre-test (t=3.11, p<.01) and the Behavior post-test (t=6.10, p<.001). A repeated measures ANOVA was conducted to see if there was a significant interaction effect between group (student versus employee) and improvement on the post test. Unfortunately, the results indicated that the students were significantly
different than the employee group \((F (1,188) = 9.14, p< .01)\), which makes it difficult to support the assertion that training had an effect. However, it is important to note that students had a decrease in scores on the post-test, while employees demonstrated a slight improvement. Therefore, while not statistically significant, it appears that training did have a marginal effect on the employee sample.

Table 19

*Means and standard deviations of behavior measures by students and employees*

<table>
<thead>
<tr>
<th></th>
<th>Students (n=109)</th>
<th>Employees (n=87)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Behavior Pre-test</td>
<td>9.35</td>
<td>2.39</td>
</tr>
<tr>
<td>Behavior Post-test</td>
<td>8.55</td>
<td>2.29</td>
</tr>
</tbody>
</table>
CHAPTER V
DISCUSSION

The following discussion will provide an overview of the findings as they relate to the hypothesized results. The discussion will be broken down into the following questions: Does test-wiseness generalize to employment settings? Can test-wiseness be effectively trained? Are there group differences in test-wiseness? Does test-wiseness training help to alleviate group differences in test-wiseness. Subsequently, theoretical explanations and methodological explanations for the results will be explored that includes the limitations of the current research and areas for future research. Finally, implications of the current research will be discussed.

Does Test-wiseness Generalize to Employment Settings?

The above findings indicate that test-wiseness is a variable that may have important implications in employment settings. First of all, there were wide ranges in subjects’ abilities to identify test-wiseness cues in the Learning measure pre-test with scores ranging from 1.00 to 7.00 out of a possible 7.00. In addition, there was a wide range in subjects’ abilities to apply test-wise skills in the Behavior pre-test measure with scores ranging from 4.00 to 17.00 out of a possible 17. Therefore, it appears that test-wiseness may have important
implications given the dramatic differences in scores. Those who are higher in
test-wiseness may have an advantage in testing situations where they are able to
apply these skills to answer questions that they do not know the answers to.

The exploratory analyses looking at job knowledge test scores also provide some interesting results in that scores on both the post-test Learning measure and the post-test Behavior measure were significantly correlated with scores on the job knowledge test. However, the pre-test scores for both the Learning and Behavior measures were not significantly correlated with the job knowledge test scores. This implies that the test-wiseness training may have been helpful in improving subjects’ scores on the employment test, which was administered three weeks after the training. When broken down by the cue dimensions, there were statistically significant correlations between the following cue dimensions and scores on the job knowledge test: longest alternative cues, most precise alternative cues, and the give-away cues.

Can Test-Wiseness Be Trained?

In determining whether test-wiseness could be trained in the current study, the results were mixed. Overall, training had a positive impact on subjects’ abilities to identify the test-wiseness cues on the 7-item Learning measure with subjects showing a significant improvement. This result supports previous research which found significant training effects (Callenbach, 1973; Dolly & Vick, 1986; Dolly & Williams, 1985; Oakland, 1972; Omvig, 1971). It also supports Langer, Wark, and Johnson (1973) who found evidence that any type of instruction on item cues resulted in higher test-wiseness scores than no training.
at all. Based on the improvements on the Learning measures, Hypothesis 1b was supported in that test-wiseness training had a significant effect on participants’ ability to identify the test taking strategies learned in training. In addition, the difference in scores resulted in an effect size of -.38.

However, when subjects were asked to try to answer questions by applying the test-wiseness skills they had just been trained on, the training appeared to have marginal effects. While there was a slight improvement after the training intervention, the difference was not significant and resulted in an effect size of only -.15. Therefore, while the findings were in the hypothesized direction, hypothesis 1c was not supported. This difference between Learning and Behavior improvements points out the need to measure various levels of training results. These results are similar to Campion and Campion’s (1987) study which found that trainees were able to learn interviewing principles but did not use what they learned. Therefore, while the subjects learned the test-wiseness cues, they were not as successful at translating this knowledge to the Behavior measure.

When the results were broken down by cue, there were significant differences between the pre and post measures for grammatical cues, sounds similar cues, and more precise alternative cues which indicates that the training was effective for these cues. However, there was a significant decrease in scores after training on unrelated alternative cues and the give away cues. There were no significant differences on the longest alternative cues and absolutes. These findings are quite interesting given the findings by Morse
(1998) who found that the use of absolutes was considerably more challenging for participants than unrelated alternatives, longest alternatives, and grammar cues. In this study, participants on the pre-test appeared to eliminate unrelated alternatives and use information from other questions (give-aways) with greater ease than the other cues and seemed to find the more precise alternative cue the most difficult (see Table 10). Therefore, like in Morse’s (1998) study, the unrelated alternatives cue was seen as easier on the pre-test, but unlike Morse, the use of give-aways was also seen as considerably easier than the other cues.

Interestingly, participants in the current study showed decreases on both the use of give-aways and eliminating unrelated alternatives, which were seen as the least difficult. This effect could be due to the training intervention actually confusing individuals and making them not trust their judgments on these cues, or alternatively the pre-test may have been easier than the post-test. Given the p-values found on the pilot sample of students, this is likely the case for the give-away cues (see Table 5), but not for the unrelated alternative cue.

When evaluating these results, it is important to consider that the items in the measures used represent the “worst case” scenario. That is, each item had a cue deliberately embedded into it. In real testing situations, it is unlikely that all items would contain such cues. Additionally, participants were informed that the items were fictional and that test-wiseness cues were included. Therefore, participants were primed to pay attention to these cues, whereas in an actual testing situation, the cues learned in training may not be as readily available to
the participants. Therefore, the improvements in test-wiseness must be interpreted with caution.

**Group Differences in Test-Wiseness?**

Another issue that was addressed in the present research was whether there were ethnic group differences in test-wiseness. This issue has been widely debated in the literature, and the results are somewhat inconclusive. In addition, various studies that have reported findings were found to be either methodologically flawed or were not specifically designed to evaluate this issue (Scruggs & Lifson, 1985). In the current research, there were no significant differences between whites and African Americans on the pre-test Learning measure or the pre-test Behavior measure. In addition to overall Behavior scores, scores on the cue dimensions were evaluated to determine if there were any significant differences between whites and African Americans. For all cues on the pre-test, there were no significant differences between whites and African Americans, however, the difference was significant at a liberal alpha level of p<.10 on the grammatical cues with African Americans scoring higher than the whites. Therefore, hypotheses 2a and 2b were not supported in that no differences existed on the pre-test measures. Unlike previous research that found ethnic group differences (Miguel, 1997; Barrett, Miguel, & Doverspike, in press; Diamond, Ayres, Fishman, & Green, 1976; Ebel, 1968; Kalechstein, et al, 1988; Dreisbach & Keogh, 1982) the current research did not reveal existing group differences. As a result, this research supports other studies that have
failed to find significant differences (Benson, Urman, & Hocevar, 1986; Yearby, 1975; Diamond, Ayres, Fishman, & Green, 1976).

Does Training Alleviate Group Differences in Test-Wiseness?

Another argument regarding race and test-wiseness involves whether different groups benefit more from test-wiseness training than others. For example, Dreisbach and Keogh (1982) discussed differential effects of training for minority groups. However, they did not directly address this issue, leaving the answer to this question unanswered. Subsequent research, however, has addressed this issue and has failed to find a significant race by test-wiseness training interaction (Benson, Urman, and Hocevar, 1986; Miguel, 1997). This provides evidence that minority group members do not necessarily benefit more from test-wiseness training.

Within the current research, the results on the Learning measure revealed that there was not a significant interaction effect with ethnicity. When looking at the behavior measure, the results were not as anticipated. While whites’ scores improved slightly, African Americans’ scores actually decreased. When exploring the interaction effects of ethnicity and training on the behavior post-test performance, the results revealed that the interaction was significant at a liberal alpha level of p<.10. Therefore, rather than diminishing group differences, test-wiseness training appeared to exacerbate the differences. It is worth noting again, however, that the pilot test revealed that the post-test was slightly more difficult than the pre-test. Therefore, the slight decrease in the African American
group may be due to measurement issues rather than the test-wiseness training intervention.

When evaluating the training effects by cues, both whites and African Americans showed significant improvements for “most precise” cues. Both whites and African Americans had significant decreases in unrelated alternatives cues and give-away cues. Whites improved significantly on grammatical cues and sounds similar cues. In contrast, African Americans improved with significance at a liberal alpha level of p<.10 for sounds similar cues and did not improve significantly on grammatical cue. In addition, scores on the cue dimensions were evaluated to determine if there were any significant differences between whites and African Americans. For all cues on the pre-test, there were no statistically significant differences between whites and African Americans, however, the difference was significant at a liberal alpha level of p<.10 on the grammatical cues with African Americans scoring higher than the whites. On the post-test, whites scored significantly higher than the African Americans on give away cues.

Therefore, in general training did not alleviate ethnic group differences and in fact resulted in greater group differences on Behavior scores overall and specifically the grammatical and give away cues.

Theoretical Explanations for the Findings

Within the above discussion, issues emerged that point out several of the limitations of the current research. Due to situational and organizational constraints, additional measures could not be included which would have been
valuable in separating out the effects of test-wiseness. The following will discuss additional issues that could help refine the impact of test-wiseness, such as cognitive ability, subjects’ motivational levels, and additional background information (e.g., social economic status, quality of the education received, etc.).

**Cognitive Ability.** In the current research, it is impossible to separate the effects of cognitive ability from test-wiseness, which would have a considerable impact on participants’ performance on the job knowledge test. Given that employment tests are often highly correlated with cognitive ability, the relationship between test-wiseness and cognitive ability is quite important. Within the test-wiseness literature, there is a debate regarding the degree to which test-wiseness correlates with general cognitive ability. Several studies have reported findings indicating that test-wiseness and cognitive ability are separate constructs. For example, Miguel (1997) found that even when the effects of general mental ability were controlled for, test-wiseness was still a significant predictor of performance on reading comprehension questions without the passages. Similarly, Crehan, Gross, Koehler, and Slakter (1978) found that test-wiseness and cognitive ability are not highly correlated. Finally, other studies have reported that test-wise individuals often score higher than those low in test-wiseness who are equal in terms of cognitive ability (Gross, 1977; Wahlstrom & Boersma, 1968).

Scruggs and Lifson (1985), however, argue that test-wiseness and cognitive ability are more closely related than others have indicated. They base their argument largely on what they feel is a lack of substantial evidence to
support the idea that test-wiseness and cognitive ability are separate constructs. Scruggs and Lifson (1985) cite findings by Anderson (1973) and Diamond and Evans (1972), which found a significant, yet moderate correlation between test-wiseness and general mental ability. Based on these findings, Scruggs and Lifson (1985) claim that test-wiseness is not a construct that “students happen to acquire by chance or serendipity, which is unrelated to intelligence, and which results in substantial fluctuations of scores in achievement tests” (p. 342).

In the current research, it is likely that cognitive ability may have impacted participants’ abilities to learn the test-wiseness cues in the allotted training program. Therefore, those who scored higher on the post-test may have been those who performed higher on the job knowledge test due to their cognitive ability. Unfortunately, the present research could not include a measure of cognitive ability due to organizational and situational constraints. Interestingly, however, the pre-test Learning and Behavior measures of test-wiseness were not significantly correlated with the job knowledge test, which suggests that test-wiseness and cognitive ability may not be as closely related as Scruggs and Lifson (1985) contend. While job knowledge tests are known to be highly correlated with measures of general cognitive ability, there are additional factors that contribute to individuals’ scores on such tests, such as prior knowledge of the material, motivation and amount of time spent studying the material. Those who were more motivated to learn the test-wiseness cues in the training program may also have been more motivated to study and prepare for the job knowledge test. Therefore, the relationship of test-wiseness and test performance could be
better refined in future research that includes a cognitive ability and a
motivational measure. It is also worth noting, however, that these significant
correlations are somewhat surprising given that the job knowledge test was
developed by professionals who have been trained in item writing and test-wise
cues and the test went through several reviews in order to ensure that test-wise
cues were not included.

Stereotype Threat. The concept of stereotype threat is another possible
explanation for the findings in this study. Stereotype threat has been offered as
an explanation for test score differences of groups such as African Americans on
cognitive ability tests and women in math (Steele, 1998; Steele & Aronson, 1995;
Wolfe & Spencer, 1996). According to this theory, members of minority groups
are often aware of stereotypes that are associated with their group. When
individuals perceive that these negative stereotypes are relevant, they feel
threatened and feel that they will be perceived of in terms of the stereotype even
if they do not believe the stereotype (Steele, 1997). The stereotype threat
theory has argued that fear or anxiety about being stereotyped interferes with
African Americans’ performance in testing situations. Research by Steele and
Aronson (1995) found that when whites and African Americans were given a
verbal ability test and were told it was a test of their intellectual ability, African
Americans performed more poorly than the whites. However, when the test was
presented as only a laboratory problem-solving exercise, whites and African
Americans performed equally well. Therefore, it has been argued that merely
changing the description of the test eliminated the performance differences
between groups. Similar findings have also been found with women’s scores on
math tests when told that there were gender differences with men performing
higher than women (Spencer, Steele, & Quinn, 1996). Stangor, Carr, and Kiang
(1998) extended this research and found that the activation of stereotypes
undermined the influence of positive feedback about performance (cited in Wolfe
and Spencer, 1996). Once stereotypes were activated, individuals’ confidence
in their abilities to perform the task were no longer relevant to their prediction of
task performance.

Steele and Aronson (1995) found that stereotype threat can be elicited
merely by asking individuals to indicate their race on test forms (Whaley, 1998).
It is thought that when individuals feel that their membership in a particular group
may be used to evaluate performance, their performance may be undermined.
The perceived effort needed to try to disprove the stereotype can be intimidating
(Steele, 1997). It is possible given the current situation that African Americans
may have felt threatened. In the exercise, they were asked to provide
demographic information, including race. In addition, civil service jobs tend to be
highly litigious. Therefore, stereotypes about group performance may have been
readily available and African Americans may have feared that these stereotypes
would be used to evaluate their performance.

Situational/Motivational Constraints. In order for individuals to benefit from
training, individuals must be prepared and motivated to learn (Goldstein, 1986).
While it was believed that subjects were considerably motivated to learn given
that the training was voluntary and was designed to help them perform well on
selection or promotion exams, it is possible that subjects may not have been very motivated to actually perform on the measures collected. This is quite possible given the fact that a total of 33 individuals were eliminated from analyses because they did not complete the biographical information or at least half of the items on all of the measures. In addition, it is possible that other motivational factors may have impacted participants' performance such as self-efficacy or locus of control. However, due to the nature of the situation, such measures were not able to be collected and would be interesting to explore in future research.

The findings related to age also indicate that motivation may have been a significant determinant of training impact. The results indicated that age was significantly correlated with the Behavior post-test but not with the pre-test. Older subjects showed considerably higher training improvements than those under forty (see Figure 4). It is quite possible that the older subjects may have had greater maturity and took the exercise much more seriously than the younger subjects. These findings are quite interesting given the fact that others have found evidence that adult subjects tended to be lacking in test-wiseness skills (Woodly, 1973; Bajtelsmit, 1975) which they felt was due to a lack of recent exposure to tests. Perhaps in the present environment, these individuals had been exposed to multiple-choice tests on a much more regular basis given their choice of profession where such tests are common for selection and promotion.

Additional Biographical Information. Additional background information of individuals would also provide some interesting insight into the effects of test-
wiseness. Exploring the socio-economic background of individuals, the quality of their education, the demographic make-up of their schools, and the extent to which they were exposed to multiple choice tests would all be helpful in determining some of the possible antecedents of test-wiseness and pin-pointing where test-wiseness training would provide the most utility.

Limitations and Methodological Explanations for Results

In addition to the constructs discussed above which would have been valuable in further determining the impact of test-wiseness, there were various methodological reasons why the results may not have been stronger or more conclusive. These issues include the absence of a control group, length of the training session, and measurement of test-wiseness.

Absence of a Control Group. The use of a control group would have also been valuable in further determining the effects of training. It is possible that mere exposure could have resulted in improved scores. However, given students' scores on the pilot study this is not likely. However, it would still have been interesting to see the differences between experimental and control group scores. As noted above, however, there were significant situational constraints which eliminated this possibility. All participants were required to receive the same treatment and the post-test provided subjects an opportunity to apply the skills they just learned.

Measurement. In future research it would be helpful to further refine the measurement used in this study. Primarily, it would be quite useful to replicate the findings and to revise the behavior measures to ensure that they are truly
comparable. Also, it would be quite helpful to increase the number of items on the measures to improve reliability. Alternatively, if an organization was agreeable, future research could use the Gibb measure of test-wiseness, which has been shown to be a useful measure. However, given the constraints of the organization in this study, the items had to reflect firefighting principles in order to be more acceptable to the participants.

Length of Training. It is also possible that improved findings would have resulted if the training had been longer or over successive sessions. Given organizational constraints, the training was limited to 45 minutes to one hour. While increasing this amount of time may have been beneficial, it is worth noting that Dolly and Vick (1986) found significant training results using a one-hour training session and Langer, Wark, and Johnson (1973) found that any training resulted in increases in test-wiseness.

Implications

While the findings in the present study were mixed in relation to the hypotheses, the findings do have bearing on how organizations should consider the issue of test-wiseness.

Within the literature, there appear to be two different theoretical approaches to the concept of test-wiseness that are not mutually exclusive (Sarnacki, 1979). The first approach views test-wiseness as a source of variance in test scores that impacts reliability and validity. According to this view, test-wiseness is a result of poor item writing and test construction, which introduces an additional source of error variance (Fagley, 1987; Diamond & Evans, 1972;
Ebel, 1972). Through utilizing test-wiseness skills, an individual is able to improve his or her score but the use of these skills also undermines the reliability and validity of the measure. Earlier studies concluded that test-wiseness has a greater impact on validity than on reliability since it represents systematic error variance that is unrelated to the criterion. Therefore, test validity is undermined because individuals’ responses may be due to their levels of test-wiseness rather than their actual knowledge (Thorndike, 1951; Stanley, 1971). Proponents of this view emphasize the need to eliminate item cues on tests in order to improve test accuracy.

"Savvy test takers know when to guess. They weed out the obvious distracters and guess at the rest, although guessing is a bad policy on most jobs. They scour the test to find items that give them clues to answering other items. They give special attention to the longest answer, knowing that it is often necessary to give more detail in the wanted answer. (I could almost have passed an Illinois driver’s test by choosing the longest answer every time (Barrett, 1998, p. 45)

The second approach views test-wiseness as a trait or characteristic of an individual. Rather than focus on psychometric issues, this viewpoint focuses on individuals’ abilities to apply test-wiseness skills. Proponents of this approach maintain that test-wiseness is best defined as an ability or trait of individuals rather than characteristics of the test. Therefore, the method to alleviate the problematic effect of test-wiseness is through training (Crehan et al, 1974). Training offers a way to ensure that all individuals taking a test possess relatively equal levels of test-wiseness. Therefore, test-wiseness should not provide an
unfair advantage to some and penalize those who are not test-wise (Sarnacki, 1979).

In considering the impact of test-wiseness, the best alternative is to consider both approaches, since neither viewpoint sufficiently covers the issues (Sarnacki, 1979). Taking the recommendations from both viewpoints would entail training test developers on test construction in general and more specifically on test-wiseness principles so that they avoid adding such secondary cues into their tests. However, this option alone may not be enough. Even tests developed by professionals have been found to contain item faults (Ellsworth, Dunnell, & Duell, 1990; Metfessal & Sax, 1958). Therefore, it would be prudent for organizations to offer test-wiseness training for candidates to ensure that all have the same opportunities. By following the recommendations of both viewpoints, organizations will be in compliance with the American Psychological Association’s Standards for Psychological Testing (1985), which states that test-taking strategies which are unrelated to test content should be explained to individuals before the test is given, especially if these strategies have been found to significantly impact test performance. This in turn would enhance the defensibility of selection and promotion procedures against attacks that test-wiseness had a significant influence.
REFERENCES


APPENDICES
APPENDIX A

HUMAN SUBJECTS APPROVAL

April 25, 2000

Ms. Susan Houston
50 Aiken Street, #371
Norwalk, CT 06351

Dear Ms. Houston:


The project qualified for Expedited Review and was approved on April 25, 2000. This approval is valid for up to one year, or until modifications are proposed in the project's protocol, whichever occurs first. In either instance, an Application for Continuing Review must be completed and submitted to the IRB.

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

Sincerely,

Gerald M. Parker
Director

GMP/gfD

Assigned Research #000404

cc:  Dr. Gerald Barrett
     Dr. Robert G. Lord

The University of Akron is an Equal Opportunity, Affirmative Action Institution.
## APPENDIX B

### BIOGRAPHICAL INFORMATION SHEET

<table>
<thead>
<tr>
<th>LAST NAME</th>
<th>FIRST NAME</th>
<th>M.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>I</td>
</tr>
<tr>
<td>J</td>
<td>K</td>
<td>L</td>
</tr>
<tr>
<td>M</td>
<td>N</td>
<td>O</td>
</tr>
<tr>
<td>P</td>
<td>Q</td>
<td>R</td>
</tr>
<tr>
<td>S</td>
<td>T</td>
<td>U</td>
</tr>
<tr>
<td>V</td>
<td>W</td>
<td>X</td>
</tr>
<tr>
<td>Y</td>
<td>Z</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SOCIAL SECURITY NUMBER</th>
<th>BIRTH DATE</th>
<th>TEST I.D. NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>000-00-0000</td>
<td>12-34-5678</td>
<td>987-65-4321</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>YEARS OF EDUCATION COMPLETED</th>
<th>YEARS OF WORK EXPERIENCE</th>
<th>RACE:</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>20</td>
<td>AMERICAN INDIAN, ALASKAN</td>
</tr>
<tr>
<td>15</td>
<td>30</td>
<td>AFRO AMERICAN, BLACK</td>
</tr>
<tr>
<td>20</td>
<td>40</td>
<td>ASIAN AMERICAN</td>
</tr>
<tr>
<td>25</td>
<td>50</td>
<td>HISPANIC</td>
</tr>
<tr>
<td>30</td>
<td>60</td>
<td>CAUCASIAN, WHITE</td>
</tr>
</tbody>
</table>

**U.S. CITIZEN:** Y N
**SEX:** M F
**VETERAN:** Y N

**IMPORTANT**
* USE #2 PENCIL
* EXAMPLE: A B C D E F
* ERASE COMPLETELY TO CHANGE
APPENDIX C

LEARNING MEASURE

1. When answering a multiple choice test item, which of the following is a clue that an alternative may be correct?
   A.* Words in the stem sound similar to words in the alternative.
   B. The alternative contains individuals– proper names.
   C. There are the same number of syllables in the stem and the alternative.

2. Which of the following words is a clue that an alternative is probably NOT correct?
   A. some
   B. never
   C.* occasionally

3. When reading the alternatives of a multiple choice test item, you may often times eliminate options because they:
   A.* are grossly unrelated to the topic.
   B. use all capital letters.
   C. contain negative adjectives.

4. When answering multiple choice items, often times the correct alternative:
   A. contains underlined words.
   B. uses the past tense of verbs.
   C.* is longer than the others.

5. When answering multiple choice items, often times the correct alternative:
   A.* is more precise than the others.
   B. contains italicized words.
   C. is a complete sentence.
6. A clue that an alternative may NOT be correct is if it contains a _________ error.
   A. pronunciation
   B. spacing
   C.* grammatical

7. When taking a multiple choice test, you may be able to determine the correct answer by:
   A. always answering A.
   B.* reading information in other items.
   C. choosing the one with the fewest letters.
APPENDIX D

BEHAVIOR MEASURES

INSTRUCTIONS TO PARTICIPANTS

The following exercise is designed to demonstrate test-taking skills. By working through this exercise, you will benefit more from the training that will follow this exercise.

The items in this exercise involve fire related issues. However, the content of the items is purely fictional. Therefore, you should not rely on any previous knowledge to answer these questions. This exercise is to demonstrate test-taking strategies, not knowledge. You are not expected to know the correct answer to these items. Instead, you should use test-taking strategies to come up with the “correct” alternative.

Please be sure to choose an answer for each item. You should mark your answers directly on the booklet. Try to answer to the best of your ability, yet do not spend a great deal of time on any one item. You should work independently. Do not discuss your responses with anyone else. Do not look at anyone else’s responses.

When you have completed this exercise, put your pen or pencil down. Please wait quietly while the remaining individuals finish the exercise.

REMEMBER:

• PLEASE ANSWER ALL OF THE QUESTIONS.
• MARK YOUR ANSWERS DIRECTLY ON THE BOOKLET.
• YOU ARE NOT EXPECTED TO KNOW THE CORRECT ANSWER TO THESE ITEMS. THESE ITEMS ARE COMPLETELY FICTIONAL.
• YOU SHOULD USE TEST-TAKING STRATEGIES TO COME UP WITH THE "CORRECT" ALTERNATIVE
• DO NOT RELY ON ANY PREVIOUS KNOWLEDGE TO ANSWER THESE QUESTIONS.
• THIS EXERCISE IS TO DEMONSTRATE TEST-TAKING STRATEGIES, NOT KNOWLEDGE.

DO NOT TURN THIS PAGE UNTIL YOU ARE INSTRUCTED
1. When Firefighter Jones is examining a trauma patient, he should be aware that the normal percentage of polydesmorpholar neutrophils found in the blood of a healthy human is: (Grossly Unrelated Alternatives)

A. 53/260  
B.* 2%  
C. 115%

2. Firefighter Jones recently attended a training seminar on firefighting equipment. He learned that slent is frequently used in the manufacturing of fire hoses. Firefighter Jones learned that the greatest advantage of using slent in fire hoses is that it: (Grammatical Cues)

A. less friction in the fire hose.  
B. the density of fire hose fibers doubles.  
C.* makes fire hoses more flexible.

3. Firefighter Jones is running a routine check on the fire engines and notices that the hydraulic ladder does not have the proper setting to ensure error free operation. Firefighter Jones determined this problem by checking the reading on the: (Alliterative Associations)

A.* hydron meter.  
B. phi gauge.  
C. fluid regulator.

4. When using Halon to fight a category 8 fire, Firefighter Jones should first ensure that: (More Precise Correct Alternative)

A. the hydraulic pressure is adequate.  
B.* the stream includes 20% cryptine.  
C. fire personnel have proper safety equipment.

5. Firefighter Jones has been battling a fire in a vacant apartment building for several hours. He notices materye is present. Firefighter Jones should: (Longer Correct Alternative)

A.* cover the length of the hose lines with kimelar.  
B. place salvo on the switches.  
C. align the flit connections.
6. During a training seminar on paramedic procedures, Firefighter Jones examined a slide of polydesmornolar neulukocyn. He learned that this substance is found in: (Correct Alternative Given Away in Other Item)
   A. urine.
   B.* blood.
   C. mucus.

7. Firefighter Jones has been assigned to repair a leaking hose. After completing the task Firefighter Jones should: (Longer Correct Alternative)
   A. test the plug.
   B.* record the repair in the log book.
   C. inform the crew.

8. Firefighter Jones has just finished his monthly review of how to properly wear oxygen tanks. Firefighter Jones learned that in order to safely ensure that one gets the correct supply of oxygen through his mask, he should: (Grammatical Cues)
   A.* screw a TSR into the tank.
   B. hooks up the TSR gauge.
   C. assembled the TSR meter.

9. When checking elevators for smoke damage in a high rise building, Firefighter Jones should: (Inclusionary Language – Absolutes)
   A. always inspect the gears in the elevator room first.
   B.* check to ensure the elevator doors work properly.
   C. never open the elevator plibon compartment.

10. Firefighter Jones has been asked to order 400 feet of new hose for the station. He should order hoses that contain: (Correct Alternative Given Away in Other Item)
    A.* slent.
    B. stagno.
    C. strayon.
11. Fire Chief Dolan is in charge of a volunteer fire department of a small township. Fire Chief Dolan receives a call of a fire late in the evening. According to the Standard Operating Procedures of the voluntary fire department, he should: *Inclusionary Language – Absolutes*

A. always contact off-duty firefighters for back-up.
B. never contact the closest municipal Fire Department for back-up.
C.* contact the scheduled reserve fire fighters for back-up.

12. Firefighter Jones has reported to a fire. It has recently snowed six inches and is 15°F Fahrenheit. While combating the fire, Firefighter Jones is operating a xylonex generator. In operating this piece of equipment he should ensure that the: *More Precise Correct Alternative*

A. battery is charged.
B. air vent is unlocked.
C.* farakat is set to 100.

13. Firefighter Jones should treat a victim with a mellite burn with: *Alliterative Associations*

A. Dalfrexis.
B. Bulofoid.
C.* Melproxin.

14. When attending a training session on the treatment of burn victims, Firefighter Jones learns that: *Inclusionary Language – Absolutes*

A.* burn victims respond well to desopin.
B. all burn victims require nexolin.
C. cryolin should never be given to burn victims.

15. Firefighter Jones recently attended a training session on fire retardant materials. He learned that a newly developed fire retardant fiber is: *Grossly Unrelated Alternatives*

A.* Quiliak.
B. wool.
C. nylon.
16. Firefighter Jones was at a conference on combat strategies for firefighters. During the conference, Firefighter Jones learned that the city with the longest average response time to a fire in 1975 was: (Grossly Unrelated Alternatives)

A. California.
B. New Mexico.
C.* Dallas.

17. Upon arriving at the scene, Firefighter Jones pulls the fire engine to where the injured fire victims are being treated by the paramedics. Firefighter Jones knows that he should: (Longer Correct Alternative)

A. park near the victims.
B. navigate around the victims.
C.* maneuver the engine between the fire and the victims.

18. Firefighter Jones is combating a category 8 fire. The most commonly used chemical to combat this type of fire is: (Correct Alternative Given Away in Other Item)

A. Milson.
B. Straynon.
C.* Halon.

19. Firefighter Jones recently attended a training session on the use of foams to suppress fires. During this training, Firefighter Jones learned that echantillon foam should be directed __________ the source of the fire. (Grammatical Cues)

A. rapidly
B.* beneath
C. bursts

20. Firefighter Jones is combating a large fire in a commercial building. The fire unit has been on the scene for six hours. The fitting pedestal was replaced two and a half hours ago. Firefighter Jones should: (More Precise Correct Alternative)

A.* turn the compound bevels two turns to the left.
B. call dispatch to inform them of the situation.
C. order the crew to respond to the dilemma.
21. Firefighter Jones is combating a fire which is being fueled by sterretania. In order to contain the fire, Firefighter Jones should use: (Alliterative Associations)

A. copranis.
B.* sterran foam.
C. nayadim.

---

BEHAVIOR MEASURE POST-TEST

INSTRUCTIONS TO PARTICIPANTS

As before, the items in this exercise involve fire related issues. However, the content of the items is purely fictional. Therefore, you should not rely on any previous knowledge to answer these questions. This exercise gives you another opportunity to use the test-taking strategies you just learned. They are not designed to evaluate your job knowledge. **You are not expected to know the correct answer to these items.** Instead, you should try and use the test-taking strategies you just learned to come up with the "correct" alternative.

Please be sure to choose an answer for each item. You should mark your answers directly on the booklet itself. Try to answer to the best of your ability, yet do not spend a great deal of time on any one item. You should work independently. Do not discuss your responses with anyone else. Do not look at anyone else's responses.

When you have completed this exercise, put your pen or pencil down. Please wait quietly while the remaining individuals finish the exercise.

1. Firefighter Jones is conducting a routine vehicle inspection before his shift. During his inspection, he notices that a rachet pin is loose. In order to resolve the problem, Firefighter Jones should: (Grammatical Cues)

A. radios the city garage and request another vehicle.
B.* attempts to fix the problem himself.
C.* notify the ranking officer and wait for his recommendation.
2. Firefighter Jones has encountered a victim with blue fingertips. He should know that this is a symptom of: *(Correct Alternative Given Away in Other Item)*

A. natorum slocum.
B. somatoform plexis.
C.* asphyxia dementia.

3. Firefighter Jones arrives on the scene of a fire and notices that the pressure valve is blocked. The next action Firefighter Jones should take is to: *(More Precise Correct Alternative)*

A. press the valve pressure reset button.
B.* change the valve pressure to 400 psi.
C. prepare to enter the burning structure.

4. While attending to a fire victim, Firefighter Jones notices that the victim had abdomalocitic tumefaction. Firefighter Jones was able to make this diagnosis by noticing that the victim had: *(Alliterative Associations)*

A. dilated pupils.
B. shortness of breath.
C.* abdominal pain.

5. When on forest jurisdiction, Firefighter Jones needs to determine his point of observation reference. To do this, he needs to use a: *(Correct Alternative Given Away in Other Item)*

A. spire.
B.* sprittle.
C. spondle.

6. Firefighter Jones has received notice that a fire has broken out in an old warehouse filled with antiques. Therefore, Firefighter Jones should be aware that __________ gas may be present. *(Correct Alternative Given Away in Other Item)*

A.* radio-bestos
B. rima-bifion
C. stagno-marflan
7. Firefighter Jones arrives on the scene of a fire and learns that the hose boom on the truck is not working properly. After resetting the hose boom, Firefighter Jones should: *(More Precise Correct Alternative)*

A.* increase the hydraulic pressure valve until it reads "350 psi".
B. bring the thermal tension up to operating level.
C. check that the engine backup generators are running.

8. Upon arriving at the scene of an apartment fire, Firefighter Jones notices the fire is giving off abnormally high levels of heat. Firefighter Jones should: *(Longer Correct Alternative)*

A. start the thermal mapping system.
B. switch the thermal range links.
C.* calculate the setting difference on the thermal inputting recorder.

9. Firefighter Jones has just returned from a training program on combat procedures. He learned that brazing is a technique used to: *(Longer Correct Alternative)*

A. overhaul fire scenes.
B.* control the spread of industrial substance fires.
C. break windows.

10. Firefighter Jones has just arrived on the scene and has taken charge of putting out the fire in a three story chemical manufacturing plant. Firefighter Jones notices that the flames coming from the building are bright blue in color. Firefighter Jones knows that this is a: *(Grossly Unrelated Alternatives)*

A.* Type IV incident.
B. situation that requires a triage center.
C. spire influencing the flames.

11. Firefighter Jones is the driver of a fire boat on the Milton River. When responding to a fire on the river bank, Firefighter Jones needs to ensure that: *(Inclusionary Language – Absolutes)*

A. the boat approaches the fire upwind.
B.* he never docks within 150 feet of the fire scene.
C. nobody is below the deck of the boat.
12. Firefighter Jones is on fire watch duty for the nearby forest jurisdiction. He should know that in order for his point of observation reference, he should place the sprittle: *(More Precise Correct Alternative)*

A. above the observation tower.
B. below the observation tower.
C.* 10 meters from the dustrop.

13. Firefighter Jones is doing cleaning detail at the fire station. When cleaning the fire station’s carillon, he should first make sure that the __________ is in place. *(Alliterative Associations)*

A.* carrin
B. loam
C. tarnit

14. Firefighter Jones is preparing to inspect the 2200SXi fire engine after a run. He must first remove __________ from the fire engine. *(Inclusionary Language – Absolutes)*

A.* all the fire masks
B. only the aluminum ladders
C. the hoses

15. Firefighter Jones has just been debriefed about using the new TSQ water pressure regulator. Firefighter Jones has learned that the TSQ regulator should be used: *(Inclusionary Language – Absolutes)*

A.* for all Type I and Type II fires.
B. when the wind speed is over 10 mph.
C. in every Code Red situation.

16. Firefighter Jones is combating a fire at an antique rug and furniture shop. He fears that the fire may be producing radio-bestos gas. Firefighter Jones should: *(Grossly Unrelated Alternatives)*

A. start running.
B.* ventilate the area.
C. remove his SCBA.
17. Firefighter Jones is combating a fire at the local high school. His commander tells him that the hullit cartridge is malfunctioning. In response, Firefighter Jones: *(Longer Correct Alternative)*

A.* reverses the crosscut processor gears.
B. loads the spare.
C. shuts it down.

18. When extricating a patient from a vehicle on the edge of a bridge, Firefighter Jones should use: *(Grammatical Cues)*

A. start pulling with a strong rope.
B.* an agit clamp.
C. grab the vehicle with a mandi bar.

19. Firefighter Jones is treating a patient for asphyxia dementia. The symptoms of asphyxia dementia include dizziness, sweating, and blue fingertips. The first thing Firefighter Jones should do is: *(Alliterative Associations)*

A. check to see if the patient's pupils are enlarged.
B.* administer the patient an asphixic muscle relaxer.
C. cover the patient with a blanket to stop shock.

20. Firefighter Jones is at the station. He has been assigned to clean the flapper valve on the pumper. After removing the flapper cap, he should: *(Grammatical Cues)*

A.* break the o-ring seal.
B. placed the fitting tube.
C. small 2 1/4 inch pliers.

21. While driving to the scene of a fire, Firefighter Jones notices that the twist anchor vessel is 90 degrees off center. When he returns to the fire house, he should: *(Grossly Unrelated Alternatives)*

A. dry the hoses.
B. log the missing axe.
C.* reset the column.

STOP HERE
AND
WAIT FOR FURTHER INSTRUCTIONS
Now that you have all completed the exercises, I would like to explain to you what these items were concerned with. You may have felt frustrated or discouraged when trying to answer these items because they were so difficult. If I was in your position of having to answer these questions, I also would been very frustrated. The reason why is because there were NO true, objectively correct answers for any of these items. However, each of the items was constructed to have cues in them which would help you to guess the correct alternative. The items were designed to familiarize you with test taking strategies that may be used when taking multiple choice tests. As such, the items were designed so that you could NOT rely on your past knowledge.

While the tests that we at Barrett and Associates, Inc. develop for selection and promotional purposes are designed to eliminate such cues, the information I am going to present is helpful in situations where you do not know the information that is being asked in a question and you have to guess. Therefore, the strategies I will discuss will help you to become a better guesser when you don’t know the correct alternative in a multiple choice item. It is important to keep in mind that these strategies are only helpful hints or rules of thumb to use. They are in no way a substitute for careful and thorough preparation.

Remember, it is to your advantage to guess. There are seven strategies that I am going to discuss today. Each of these strategies will help to make you a more effective guesser. These seven strategies are:

- **CHOOSE WORDS IN THE STEM THAT SOUND LIKE ONE OF THE ALTERNATIVES.**
- **AVOID UNREALISTIC ALTERNATIVES.**
- **LOOK FOR KEY WORDS WHICH SUGGEST THAT AN ALTERNATIVE IS INCORRECT.**
- **SELECT LONGER ALTERNATIVES.**
- **CHOOSE MORE PRECISE ALTERNATIVES.**
• LOOK FOR GRAMMATICAL CLUES.
• USE INFORMATION FROM OTHER ITEMS TO HELP ANSWER QUESTIONS.

We will now go over each of these seven strategies in more detail.

SIMILAR SOUNDING ALTERNATIVES

Sometimes you will be able to identify a correct alternative because it sounds similar to a word in the stem of the question. For example, an item from the exercise you just completed stated:

Firefighter Jones should treat a victim with a Mellite burn with:

A. Dalfrexis.
B. Bulofoid.
C. Melproxin.

In this item, the stem contains the word “mellite”. Given that there is no such thing as a mellite burn, there is no real correct answer to this item. However, the alternative “Melproxin” sounds most like the word “mellite” in the stem. Therefore, if you had to guess, it is likely that “C” would be the correct alternative.

For items where you do not know the correct answer and have to guess, often times the alternative which sounds similar to words or phrases in the stem is the correct one.

UNRELATED ALTERNATIVES

Sometimes, the correct alternative to an item is determined by eliminating other alternatives. Specifically, some alternatives may be grossly unrelated to the topic of the item. These alternatives can then be eliminated which improves your chances of guessing correctly. For example, an item from the exercise you just completed stated:

Firefighter Jones was at a conference on combat strategies for firefighters. During the conference, Firefighter Jones learned that the city with the longest average response time to a fire in 1975 was:

A. California.
B. New Mexico.
C. Dallas.
In this item, the stem asks for a city. However, two of the alternatives are states. Therefore, the alternatives “A” and “B” can be eliminated since they are not cities. This leaves “C” as the logically correct alternative.

For items where you do not know the correct answer and have to guess, often times you can eliminate alternatives which are grossly unrelated to the information in the stem.

ABSOLUTES

Yet another strategy involves avoiding certain “key words”, or “absolutes” within alternatives. Such words often imply that an alternative is incorrect because these words are very broad and difficult to defend. Therefore, in avoiding these words, you may be able to eliminate one or more alternatives. You may then be able to guess among a smaller group of alternatives. Often, alternatives that contain the following words should be avoided:

ALWAYS
ALL
NONE
NEVER
EVERYONE
NOTHING
ONLY
NOBODY

Alternatives which contain words like these are difficult because rarely do we come across situations where something is absolute or true 100% of the time. Usually, we can come up with exceptions to the rule. Therefore, saying that something happens “always” or “never” is problematic because we can usually come up with an exception which implies that this alternative is incorrect. Even if you can’t come up with an exception yourself, there may still be a particular situation which violates this alternative. Therefore, when you run across alternatives which contain words such as those listed above, you may be fairly safe in assuming that you can eliminate them. For example, an item from the exercise you just completed stated:

When attending a training session on the treatment of burn victims, Firefighter Jones learns that:

A. burn victims respond well to desopin.
B. all burn victims require nexolin.
C. cryolin should never be given to burn victims.
In this item, alternatives “B” and “C” contain absolute words. Therefore, these alternatives can be eliminated. This makes “A” the logically correct alternative. Remember, however, that this is merely a guessing strategy. This strategy will not work on all occasions. For items where you do not know the correct answer and have to guess, you may be able to eliminate alternatives that contain absolute words.

**LONGER CORRECT ALTERNATIVES**

With some items, the correct alternative is different in form than the other alternatives. In particular, the correct alternative may often be the longest alternative. Alternatives which are longer are often correct because the item writer wanted to make sure that all relevant or important information was included. For example, an item from the exercise you just completed stated:

> Upon arriving at the scene, Firefighter Jones pulls the fire engine to where the injured fire victims are being treated by the paramedics. Firefighter Jones knows that he should:

A. park near the victims.
B. navigate around the victims.
C. maneuver the engine between the fire and the victims.

In this item, the alternative “C” is the longest. While it is not necessarily true that the longest alternative is the correct one, often times it is. Given that the item above is fictional, there is no real correct answer. However, alternative “C” is the longest. Therefore, if you had to guess, it is likely that “C” would be the correct alternative.

For items where you do not know the correct answer and have to guess, often times the alternative which is the longest is correct.

**MORE PRECISE ALTERNATIVE**

As in the situation where the correct alternative is often the longest one, the most precise alternative is also often the correct answer. Alternatives which contain more detail or are more precise are often correct because the item writer wanted to make sure that all relevant or important information was included. For example, an item from the exercise you just completed stated:
When using Halon to fight a category 8 fire, Firefighter Jones should first ensure that:

A. the hydraulic pressure is adequate.
B. **the stream includes 20% cryptine.**
C. fire personnel have proper safety equipment.

In this item, alternative “B” contains more detail and is more precise. The other two alternatives, while they are plausible, are more vague. Therefore, if you have to guess, you may be more successful by choosing an alternative that has more detail.

For items where you do not know the correct answer and have to guess, often times the alternative which is the most precise is the correct one.

**GRAMMATICAL CUES**

Some items may contain grammatical errors or inconsistencies which can help indicate the correct alternative. For example, the stem may end with the word “an”. Usually, the word “an” indicates that the following word begins with a vowel. If an alternative begins with a consonant, this may imply that the alternative is not correct. Alternatively, the verb tense may be different in the stem than in the alternatives. This difference in verb tense may indicate that an alternative is incorrect and should be avoided. For example, an item from the exercise you just completed stated:

Firefighter Jones has just finished his monthly review of how to properly wear oxygen tanks. Firefighter Jones learned that in order to safely ensure that one gets the correct supply of oxygen through his mask, **he should**:

A. **screw** a TSR into the tank.
B. **hooks** up the TSR gauge.
C. **assembled** the TSR meter.

In this item, the last words in the stem read “he should”. The first words of alternatives “B” and “C” do not flow because they are not in the same verb tense. The phrases “he should hooks” and “he should assembled” are not grammatically correct. Therefore, alternative “A” would be a good guess because it is grammatically correct.

For items where you do not know the correct answer and have to guess, often times the alternatives which are grammatically incorrect should be avoided.
GIVE - AWAYS

Sometimes you may find clues or information in other questions within the test that may help you answer a particular question. By carefully reading each item, you may discover that some items contain similar information. In these situations, you may be able to find the answer to one item in a different item in the test. For example, items from the exercise you just completed stated:

During a training seminar on paramedic procedures, Firefighter Jones examined a slide of polydesmopholar neulukocyn. He learned that this substance is found in:

A. urine.
B. blood.
C. mucus.

When Firefighter Jones is examining a trauma patient, he should be aware that the normal percentage of polydesmopholar neulukocyn found in the blood of a healthy human is:

A. 53/260
B. 2%
C. 115%

The answer to the first question is contained in the second item. The second item contains the phrase "polydesmopholar neulukocyn found in the blood". This phrase gives away the answer to the first question which asks where polydesmopholar neulukocyn is found. Therefore, the correct answer to the first item would logically be alternative “B”.

Therefore, when you are unable to answer a question, it may be a good idea to look over the other items on the test to see whether there are any clues within them which may help you answer other items.

SUMMARY

In conclusion, when taking a multiple choice test, you may run across items where you do not know the correct answer. In such situations it is usually to your advantage to guess. Therefore, it is helpful to know how to guess more effectively. The strategies we went over today are designed to help you become a better guesser in these situations. However, it is always best to be well prepared so that you do not have to guess since these strategies are not fool proof. Test developers are aware of these strategies and make efforts to
eliminate them. Nevertheless, it is possible that being aware of them may help you in a testing situation if you need to guess. To review, the seven strategies we discussed today included:

- WORDS IN THE STEM THAT SOUND LIKE ONE OF THE ALTERNATIVES.
- UNREALISTIC ALTERNATIVES.
- KEY WORDS WHICH SUGGEST THAT AN ALTERNATIVE IS INCORRECT.
- LONGER CORRECT ANSWERS.
- MORE PRECISE CORRECT ANSWERS.
- GRAMMATICAL CLUES.
- GIVE AWAYS FROM OTHER ITEMS.