Navy Personnel Research, Studies, and Technology Division
Bureau of Naval Personnel (NPRST/BUPERS-1)

Millington, TN 38055-1000 **NPRST-TR-08-2** **August 2008**

Evaluation of the Fake Resistance of a Forced-choice Paired-comparison Computer Adaptive Personality Measure

Christina M. Underhill
Ronald M. Bearden
Hubert T. Chen

NPRST
research at work

NPRST-TR-08-2
August 2008

Evaluation of the Fake Resistance of a Forced-choice Paired-comparison Computer Adaptive Personality Measure

Christina M. Underhill
Ronald M. Bearden
Hubert T. Chen

Reviewed and Approved by
Jacqueline A. Mottern, Ph.D.

Released by
David L. Alderton, Ph.D.

Navy Personnel Research, Studies, and Technology (NPRST/BUPERS-1)
Bureau of Naval Personnel
5720 Integrity Dr.
Millington, TN 38055-1000
www.nprst.navy.mil

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
08-31-2008	Technical Report	July 2005 - August 2006

4. TITLE AND SUBTITLE

Evaluation of the Fake Resistance of a Forced-choice Paired-comparison Computer Adaptive Personality Measure

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Christina M. Underhill
Ronald M. Bearden
Hubert T. Chen

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Navy Personnel Research, Studies, and Technology (NPRST/BUPERS-1)
Bureau of Naval Personnel
5720 Integrity Dr.
Millington, TN 38055-1000

8. PERFORMING ORGANIZATION REPORT NUMBER

NPRST-TR-08-2

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Navy Personnel Research, Studies, and Technology (NPRST/BUPERS-1)
Bureau of Naval Personnel
5720 Integrity Dr.
Millington, TN 38055-1000

10. SPONSOR/MONITOR'S ACRONYM(S)

NPRST

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

NPRST-TR-08-2

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report documents research that supports the use of the Navy Computer Adaptive Personality Scales (NCAPS) as a fake-resistant alternative when compared with other personality measures using a Likert-scale format. NCAPS is a computer adaptive personality measure being developed and validated for use in the selection and classification of Sailors for entry level Navy enlisted jobs. NCAPS uses a cutting-edge technological approach to personality measurement which is designed to mitigate many problems that plague traditional instruments. Specifically, traditional instruments use straight-forward Likert rating scales where respondents specify their level of agreement to a statement. Moreover, such instruments generally contain sets of homogeneous items with a transparent content, which makes them relative easy to fake (good or bad) and subject to social desirability bias (making oneself look). To minimize these problems, NCAPS developed a paired-comparison forced-choice item format, uses a complex item response theory (IRT) adaptive selection and scoring algorithm, and intersperses item content. Results were striking. There were no significant mean differences between honest and faking scores on any of the 10 personality traits measured by the adaptive test. There were however, significant mean differences between honest and faking scores on all 10 traits measured by the Likert-scale NCAPS. Simply stated,

15. SUBJECT TERMS

personality assessment, response distortion, Likert-scale format, adaptive assessment in personality

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Genni Arledge
UNCLASS	UNCLASS	UNCLASS	UNCLASS	37	19b. TELEPHONE NUMBER *(Include area code)* 901-874-2115 (882)

Foreword

This report documents research that supports the use of the Navy Computer Adaptive Personality Scales (NCAPS) as a fake-resistant alternative when compared with other personality measures using a Likert-scale format. NCAPS is a computer adaptive personality measure being developed and validated for use in the selection and classification of Sailors for entry level Navy enlisted jobs. The program is designed to replace the current classification algorithm with a more flexible and accurate one, de-emphasize the almost exclusive focus on mental ability by including personality and interest measures in making classification decisions, and to better understand "Sailorization" process and how it contributes to attrition. Collectively, these efforts are transforming and modernizing enlisted classification by making it applicant-centric while improving job satisfaction and performance, reducing attrition, and increasing continuation behavior.

NCAPS uses a cutting-edge technological approach to personality measurement which is designed to mitigate many problems that plague traditional instruments. Specifically, traditional instruments use straight-forward Likert rating scales where respondents specify their level of agreement to a statement. Moreover, such instruments generally contain sets of homogeneous items with a transparent content, which makes them relative easy to fake (good or bad) and subject to social desirability bias (making oneself look). To minimize these problems, NCAPS developed a paired-comparison forced-choice item format, uses a complex item response theory (IRT) adaptive selection and scoring algorithm, and intersperses item content. The complexity and novelty of the design constraints requires a series of interrelated research projects. This report covers how the adaptive paired-comparison forced-choice format used by NCAPS is less resistant to response distortion when compared to a Likert-scale NCAPS format.

The research was sponsored by the Office of Naval Research (Code 34) and funded under PE 0602236N and PE 0603236N.

DAVID L. ALDERTON, Ph.D.
Director

Executive Summary

Traditionally and currently, Navy recruits are selected, classified, and assigned to training and career paths based on a cognitive ability test known as the Armed Services Vocational Aptitude Battery (ASVAB). This is true even though we know that cognitive ability alone is not an adequate predictor for all of the outcomes currently important to the Navy, such as good citizenship, teamwork propensity, job satisfaction, job performance, and continuation behavior. In particular, it has long been known that personality measures can dramatically improve the predication of non-training outcomes. This shortfall served as the impetus for developing the Navy Computer Adaptive Personality Scales (NCAPS). NCAPS was designed to serve as a non-cognitive complement to the ASVAB.

In established personality instruments, Likert-scales are universally used and these are vulnerable to social desirability bias, particularly when instruments are used for high-stakes decision making (e.g., offering employment). To address this concern, NCAPS measures personality utilizing a computer-adaptive, paired-comparison forced-choice item format. The research described in this report provides evidence that the computer adaptive methodology and item formats in NCAPS are fake-resistant when compared with other personality measures using a Likert-scale format.

Participants in this study were recruited from introductory psychology courses and several online wellness courses at an urban university. A total of 158 students participated. Respondents were asked to take either the adaptive version or non-adaptive version of NCAPS, twice. They first answered the questions honestly, then answered the items a second time purposely trying to inflate their scores (i.e., present themselves as the ideal employee).

Results were striking. There were no significant mean differences between honest and faking scores on any of the 10 personality traits measured by the adaptive test. There were however, significant mean differences between honest and faking scores on all 10 traits measured by the Likert-scale NCAPS. Simply stated, participants were not able to intentionally distort their personality scores when taking the adaptive paired-comparison NCAPS. As has been demonstrated before, on the traditional Likert-scale version, participants were easily able to significantly distort their scores, on every one of the 10 personality scales. Moreover, on the traditional Likert-scale version of NCAPS, participants higher in cognitive ability and reading ability were able to produce higher fakability scores. Higher intelligence and reading scores had no effect on a participant's ability to fake the adaptive, paired-comparison version.

In summary, these results support the notion that, not only is the adaptive, paired-comparison version of NCAPS fake-resistant in general, but this is true even among those with of high intelligence and reading ability. Therefore, the adaptive paired-comparison NCAPS is very likely to provide scores close to the true trait scores for an individual even under high-stakes testing conditions.

Contents

Evaluation of the Fake Resistance of a Forced-choice Paired-comparison Computer Adaptive Personality Measure **1**

Personality Measures in the Navy **1**

Initial NCAPS Validation 4

Impact of Faking 5

Paired-comparison Formats 6

Measurement of Faking 6

Predictors of Faking 8

Procedures **8**

Overview 8

Participants 9

Measures and Materials 9

Design and Study Procedures 10

Data Scoring 11

Data Integrity 11

Group Differences 11

Results **11**

Cognitive Ability and Reading Ability. 13

Discussion **14**

References **15**

Appendix **A-0**

List of Tables

1. Traditional NCAPS item presentation vs. Adaptive NCAPS item presentation 4
2. Instructions to Participants 10
3. Mean scores and standardized mean differences 12
4. Regression model statistics 13

Evaluation of the Fake Resistance of a Forced-choice Paired-comparison Computer Adaptive Personality Measure

Unlike the Armed Services Vocational Aptitude Battery (ASVAB) or other tests of intellectual ability, generally there are no right or wrong answers on personality tests (e.g., extroversion, openness to experience). However, there are socially desirable traits and there are characteristics that are preferred by employers. Because these are generally known (i.e., socially desirable and employer preference), faking on personality tests in employment settings is a common problem. The purpose of this research project is to provide evidence regarding the fake resistance of the Navy Computer Adaptive Personality Scales (NCAPS). NCAPS is a forced-choice paired-comparison computer-adaptive personality measure developed at the Navy Personnel Research, Studies, and Technology (NPRST) division, which is the Navy's personnel research laboratory. This study compares the fake resistance of two forms of NCAPS, the adaptive paired-comparison version and the non-adaptive Likert-scale version. This is the first study to evaluate the extent to which participants can deliberately elevate their personality scores on this adaptive NCAPS measure.

Participants in this study were asked to take either the adaptive version or non-adaptive version of NCAPS, twice. The first time the participants were instructed to take the measure honestly. The second time they were instructed to deliberately fake to make the best impression possible for obtaining a job. Differences in individual personality scores from the honest and fake instructions were compared between the adaptive paired-comparison form and the Likert-scale form. Faking or response distortion was operationally defined as an increase in trait scores from the honest condition to the fake condition (e.g., a participant who says they are more dependable in the faked version than in the honest version). It was hypothesized that participants would have more difficulty purposely inflating their scores on the paired-comparison adaptive version of NCAPS than on the Likert-scale version.

Personality Measures in the Navy

To enlist in the Navy, applicants must meet the minimum requirements on the Armed Services Vocational Aptitude Battery (ASVAB), a test battery that assesses performance in reading, mathematics, and general science, as well as basic knowledge about electronics, automotive and shop information, and mechanical systems. A classifier[1] uses combinations of ASVAB subtest scores and identifies which technical training schools the applicant is qualified for and likely to pass, this list is then compared to a list of available jobs. The classifier attempts to interest the applicant in one of the jobs in a short interview. At the conclusion of this meeting, the classifier and applicant come to an agreement and a contract is signed guaranteeing the technical training school, basic training start date, and any special addendums (e.g., an enlistment

[1] In military entrance processing, duties are separated between the recruiter, who "sells" the Navy to the applicant, and the classifier, who sells the specific job, training, and start date to the applicant.

bonus). However, as Borman, Hedge, Ferstl, Kaufman, Farmer, and Bearden (2003) discussed in their review of selection and classification, individuals are more complex and multidimensional than the cognitive abilities assessed by the ASVAB. Beyond cognitive abilities, individuals possess a variety of preferences, interests, and personal characteristics that are predictive of good citizenship, teamwork, job satisfaction, job performance, and continuation behavior. The current Navy classification process does not utilize any non-cognitive information for job placement (except for casually stated preferences).

The goal of researchers in personnel selection and classification is to develop measures that predict job performance and/or job tenure. Measures given to job applicants need to assess the knowledge, skills, and abilities necessary for successful performance in a particular job, ideally without producing adverse impact (large mean differences) for racial, ethnic, or gender groups. Cognitive ability is the single best predictor of both training and job performance (Hunter & Hunter, 1984; Godfriedson, 1986; Ree, Earles, & Teachout, 1994; Schmidt & Hunter, 1998). However, studies by Borman, White, and Dorsey and by Borman, White, Pulakos, and Oppler (as cited in Ferstl, Schneider, Hedge, Houston, Borman, & Farmer, 2003), found that in certain domains of job performance the variance accounted for can increase substantially when personality measures are used in conjunction with cognitive ability measures (see also, McHenry, Hough, Toquam, Hanson, & Ashworth, 1990).

Just as cognitive ability alone cannot predict who will be successful in all critical performance domains; cognitive ability alone is not sufficient for predicting whether a person will fit well with his or her organization and remain on the job. Employers generally want employees who not only perform well on the job but also remain on the job. Research has shown that one's personality, motivation, and interest substantially help predict turnover, retention, and job performance (Borman et al., 2003). In general, cognitive ability predicts knowledge components of job performance, whereas personality variables are better at predicting motivational components of performance (McCloy, Campbell, & Cudeck, 1994), which influence turnover and retention.

Many studies have found that measuring personality variables greatly enhances our ability to predict who will perform successfully across a variety of jobs in civilian and military settings. For instance, conscientiousness is one of the best personality traits for predicting performance across a variety of jobs. By adding a measure of conscientiousness, an additional 18 percent of variation in on-the-job performance can be explained. In fact, an investigation with military participants found that measuring emotional stability accounted for an additional 38 percent of job performance variance (see Ferstl et al., 2003).

In short, research evidence indicates that the assessment of personality is a very promising approach to achieve greater operational and economic efficiencies in the Navy, yet personality tests are still not incorporated into Navy selection or classification. There are many historical and practical reasons for this. Most personality tests were designed to detect psychopathology and not to predict performance in the armed services. While there have been a few large-scale studies of personality and job performance, most are limited to small groups. Most personality tests are too long and cumbersome to be delivered efficiently. Perhaps most importantly, personality tests

have not been widely validated against actual on-the-job performance across the many different occupations in the Navy. However, the single most important reason that personality tests are not used for operational selection and classification decisions is that traditional personality instruments are relatively easy to fake to make the applicant look better than he or she actually is.

The Navy Computer Adaptive Personality Scales (NCAPS) was developed to provide the Navy with an efficient measure of personality traits on which to better classify Navy recruits. NCAPS is an adaptive measure that uses item response theory (IRT) methodology to modify item presentation based on test takers' responses, which in turn decreases the number of items presented, and reduces testing time, while improving the accuracy of test scores. NCAPS presents items in pairs, and responders are forced to choose one or the other. This forced choice format has been shown to be more resistant to faking other forms of response distortion (Jackson, Wroblewski, & Ashton, 2000; Martin, Bowen, & Hunt, 2002). Personality constructs measured by NCAPS were chosen based on their relevance and criticality to job performance in many entry level Navy enlisted jobs. For a more detailed description of the process that identified the 10 traits measured by NCAPS, see Houston, Borman, Farmer, and Bearden (2005). The current study assesses the fake resistance of NCAPS.

The main principle behind adaptive testing used in employee selection is that the person's prior responses to test items are used to determine the next test item to present. All adaptive test item selection algorithms use item difficulty to determine the next item in a sequence. If a participant responds correctly to an item, then he or she is presented with a more difficult item. If the participant responds incorrectly, he or she is presented with a less difficult item. Items are presented until the participant consistently answers items correctly at a specific level of difficulty or other statistical criteria are met (Bartram, 1993; Wainer, 2000). In personality testing, "difficulty" does not take on the standard meaning in an ability test; instead a difficult item is one that is higher on the trait of interest (e.g., on a measure of extraversion, "I like parties" would be considered a higher trait item than "I like libraries").

In many testing environments, including military personnel testing, there is a limited amount of time available for assessment. Therefore the purpose of computer-adaptive testing is to present items that are informative about the test taker and to maximize the precision of measurement in a limited amount of testing time. For example, on a standard cognitive ability test a high ability person will receive the same easy items as everyone else, yet they will contributed little to no information about his or her actual ability. Only the more difficult items will provide information about the person's actual ability. By using adaptive testing methods, the high ability person will not be administered the easy items. Similarly, a low ability person will not receive the more difficult items. But only administering items that are informative of the person's ability, the number of test items can be greatly reduced along with the administration time (Wainer & Mislevy, 2000). A similar approach is taken when measuring a person's trait level using NCAPS.

Computer-adaptive tests developed since the low cost and easy availability of high-powered computers (e.g., Graduate Record Examination [GRE] and American College Test [ACT]), test job knowledge and cognitive ability. Computer-adaptive technology

(CAT) has not yet been applied to the measurement of personality; therefore, there is very little research regarding computer-adaptive personality testing (Ferstl et al., 2003; Wainer, Dorans, Green, Mislevy, Steinberg, & Thissen, 2000). Prior to NCAPS, there have been no reports of a functional computer-adaptive personality measure in the literature. Again, when measuring personality as opposed to measuring cognitive ability, there is no right or wrong answer or degree of difficulty. Items on a personality test are differentiated by how strongly each statement represents a particular personality trait. For example, a statement representing someone with low achievement is, "I only take on projects that I expect will be easy to complete." A statement representing someone with high achievement is, "I usually set difficult goals for myself." For a complete description of item development and trait scaling for NCAPS, see Ferstl et al. (2003) and Houston et al. (2005).

NCAPS is a paired-comparison forced-choice measure. Several methods of computer-adaptive testing were explored for this endeavor and a statistical method refined by Stark and Drasgow (2002) was selected (see also Houston, et al., 2005). Test takers are presented two statements representing two different levels of a trait and asked to choose which of the two statements is most descriptive of him or her. The response causes the program to branch to a greater or lesser level for that particular trait. Traditional and Adaptive presentations are depicted below in Table 1.

Table 1
Traditional NCAPS item presentation vs. Adaptive NCAPS item presentation

Traditional Item Presentation	Adaptive Item Presentation*
I always do the work that is expected of me A. This describes me all of the time B. This describes me most of the time C. This describes me some of the time D. This describes me rarely E. This doesn't describe me	I always do the work that is expected of me (trait value = 3) I like to set goals that force me to perform at a level higher than what I've done in the past (trait value = 5)

*The adaptive item presentation asks the test taker to choose one of the two statements presented. The trait value is provided for your reference (the test taker would not see the trait values). The adaptive process is explained in more detail below.

Initial NCAPS Validation

Initial tests of the NCAPS program have been very successful. Pilot testing has indicated that NCAPS has good construct validity, demonstrating that the items are measuring their intended constructs. NCAPS has been tested on small samples of college students and first-term enlisted Sailors. Results of the tests with college students found that ACT scores, a cognitive ability measure, were not related to the personality traits. However, certain personality traits such as achievement motivation were significantly related to classroom and college performance. This finding for incremental

validity is concordant with the established literature and further demonstrates that cognitive abilities are not related to personality, and that personality traits usefully supplement cognitive ability in predicting training performance (Underhill, 2004). Testing of first-term enlisted Sailors showed that various personality traits as measured by NCAPS are significantly related to different aspects of job performance as indicated by supervisor ratings.[2]

Impact of Faking

While research has shown that personality measures can increase the performance prediction above what can be predicted by cognitive ability alone (Schmidt & Hunter, 1998), personality measures are the most susceptible to faking and other forms of response distortion (Borman et al., 2003). "Faking good" is a participant's inflation of responses on a measure to make them appear more favorable. The identification of people who fake or distort their responses on personality measures is a popular and longstanding topic for psychologists and human resource managers. Research has shown that when a person does not accurately respond and they inflate their scores, they have a better chance of getting hired for the job (Mueller-Hanson, Heggestad, & Thornton, 2003; Rosse, Stecher, Miller, & Levin, 1998). A review of studies by Hough (1998) revealed that intentional distortion has little effect on the criterion validities of personality measures. Nevertheless, faking still concerns practitioners because more flagrant distorters have been shown to be more likely to be selected in a top-down selection process.

Mueller-Hanson et al. (2003) examined faking in an incentive group and its impact on selection. The authors found that when there is a smaller selection ratio, larger numbers of people from the incentive group would be hired over people in the honest groups. Rosse et al. (1998) also found that there was an overrepresentation of identified fakers in the top 5 percent of job applicants. Both studies found that as the selection ratios decrease, more fakers than honest respondents are hired, but when the selection ratio increases and more people are hired for the job, then the numbers of potential fakers and honest responders hired evens out (Mueller-Hanson et al., 2003; Rosse et al., 1998). Since there is a potential for hiring more fakers combined with the lack of solid and prevalent evidence of faking on job performance of actual applicants, it is important to create measures that reduce a person's ability to fake. Such measures create a more even playing field, because even if an applicant had the ability and/or motivation to purposely increase their scores, they would have a difficult time doing so.

[2] At the time this report was originally written, this was the extent of the available data on NCAPS. Unfortunately, the lead author moved to another agency and the manuscript languished. Instead of updating some sections of the document and having to coordinate with a long-departed author, it was decided to keep the document as-is and footnote significant changes. As of the summer of 2008, well over 22,000 Sailors have taken NCAPS and there is a much more substantial basis for its validity than when the report was originally written.

Paired-comparison Formats

A primary goal of the NCAPS design was to inhibit the ability to fake. NCAPS was designed in a forced-choice paired-comparison format, which for other measures has been shown to reduce response distortion (Jackson, Wroblewski, & Ashton, 2000; Martin, Bowen, & Hunt, 2002). Jackson et al. (2000) administered an integrity test in a single stimulus (i.e., one statement with Likert-scale options) and a forced-choice format with four statements per presentation. Participants were assigned to one format or the other and asked to take the form twice, once honestly and once as if they were a job applicant. They found that participants could increase their scores on both forms under the job applicant instructions, yet there were smaller increases in mean scores on the forced-choice format indicating that it was more difficult to fake. Not only did they find that the forced-choice version was more difficult to fake, they also found that scores from this measure were predictive of behavior in the directed faking condition, whereas the scores from the Likert scale were not. The forced-choice format therefore achieved two goals: it reduced the magnitude of faking and retained criterion-related validity.

Martin et al. (2002) also compared the fake resistance of forced-choice and Likert-scale formats. In their experiment, participants were assigned to either a fake or honest condition and asked to take both an ipsative (i.e., forced-choice) and normative (i.e., Likert-scale) form of the Occupational Personality Questionnaire. Faking was operationalized by how close participants were able to match their responses on the measures to what they thought were the ideal characteristics of a junior manager. A closer distance between their score and their ideal rating indicated a greater ability to fake. Participants in the honest condition had greater distances or discrepancy between their scores and what they thought were ideal traits because they were not asked to fake toward their ideal. Participants in the fake condition had much smaller distances, indicating that they were able to match their scores more closely. The prominent finding in this study was the difference between the scores on the forced-choice and Likert-scale among the participants in the faking condition. Results indicated that people had a much more difficult time in distorting their response to match their ideal on the forced-choice format than on the Likert-scale measure.

Measurement of Faking

Traditional designs of faking studies have compared differences in group means and standard deviations between applicant and incumbent groups or experimental groups instructed to either “fake good” or “be honest.” In applicant versus incumbent groups, it is assumed that applicants are more motivated to distort their responses to appear more favorable in order to be selected for a job. It is also assumed that job incumbents would respond to measures honestly because they already have their job and have little reason to distort their response. This response distortion has been measured by increases in mean scores of the applicant group over those of the incumbent groups. Rosse et al. (1998) found that applicants had higher personality scores on more favorable traits (e.g., agreeableness) and lower scores on less favorable traits (e.g., neuroticism) than incumbents. Research comparing experimentally manipulated groups (where one group is given an incentive and is directed to distort their response, and another group is asked

to respond honestly), has also found significant increases in scores from the incentive (faking) groups over the honest group. Mueller-Hanson et al. (2003) also found that the incentive group, when compared to an honest group, scored significantly higher on a measure of achievement.

Other research has compared within-subject differences between responses in an honest and fake condition. There have been differences in results of the sensitivity of the statistics used to indicate a person's ability to fake. In 1986, Lautenschlager described four within subject measures for the assessment of individual differences in faking. Two of the measures were previously reported in Gordon and Cross (1978), as referenced in Lautenschlager (1986) review of the literature, on methods to detect faking on self-report measures. Gordon and Gross concluded that the overall difference in mean scores under an honest and fake condition as well as the variance of these difference scores were useful methods to detect faking. Lautenschlager compared these two methods and proposed two additional measures to detect faking (a) correlation of scores from the honest and fake conditions indicating the consistency of a subject's responses under the different response conditions, and (b) the within-subject variance of the differences in item responses from honest to fake condition.

Mersman and Shultz (1998) followed Lautenschlager's recommendation for using these measures. They used three indices of faking ability: within subject correlations between honest and faking scores, mean differences between honest and faking scores, and within subject variance of the differences in item responses between honest and faking conditions. The three faking indices did not produce the same results in their analyses. The correlation index showed some variability in responding, but participants generally responded consistently from the honest to fake condition. The within-subject variance of the differences index provided "insignificant and erratic correlations" with the factors they used to explain individual differences in faking ability (p. 225). The one index of faking that showed significant differences between the honest and faking scores was the mean difference. The *t*-tests on the differences between means showed that participants could significantly increase their scores from the honest condition to the fake condition.

Zickar, Gibby, and Robie (2004) proposed a new method to identify fakers on personality measures, mixed model item response theory (MM-IRT). Zickar et al. purported that a problem with previous research is the assumption that respondents in experimentally manipulated groups respond like they are asked or that all applicants are fakers. Zickar et al. used MM-IRT to investigate the number of groups and subgroups that can be reliably identified from two datasets based on response patterns. One dataset consisted of applicant and incumbent responses to the Personal Preference Inventory and the other dataset consisted of an experimentally induced faking study in which the participants took the Army's ABLE scale. MM-IRT combines latent class analysis that can identify classes of individuals (e.g., fakers and non-fakers) with IRT that can identify, based on item responding patterns, groups within the fakers and non-fakers that don't respond similarly to their class.

Results of Zickar et al.'s (2004) analyses showed that not every respondent distorts their responses the same way or to the same extent on every personality scale. Some applicants (commonly assumed to be faking) appeared to be responding honestly. They

also found that some incumbents respond as if they are faking even though they had no motivation to do so. The overall conclusion was that the research on faking that uses applicant groups as fakers and incumbent groups as honest responders is not accurate. Their results also indicate that experimental manipulations to induce faking or honest behavior did not produce consistent response patterns within each condition. Zickar et al. (2004) reported a "sizeable percentage" of participants in the honest condition who were placed in the faking class as well as participants in the faking conditions who were placed in the honest class based on their MM-IRT analyses. The researchers suggested that these differences could be "ascribed to a variety of factors, such as ability to fake, miscomprehension of the instructions, and the level of self-insight." (p. 186). They also found that the identified fakers differed in who faked what personality scales. The fakers faked more on some constructs than on others. Zickar et al. (2004) hypothesized that people may believe that certain constructs are more important than others and/or that some personality scales are easier to fake or more socially desirable than others.

Predictors of Faking

Mersman and Shultz (1998) looked at individual differences in ability to fake a measure of the Big Five. They found that neither social desirability, impression management, nor conscientiousness could explain an individual's ability to fake or increase their scores on their measure. McFarland and Ryan (2000) also investigated personality constructs related to faking or the differences in participant's scores between an honest and fake condition. They found that participants scoring high on integrity were least likely to purposely increase their scores on extroversion, agreeableness, and conscientiousness, perhaps because their scores were higher on these scales to begin with. They also found that conscientiousness was related to faking. Results indicated that more conscientious people faked less than those lower on conscientiousness. The current NCAPS study will also examine the relationship between an individual's honest score on achievement motivation, which most closely mirrors conscientiousness, and honest integrity scores with their ability to fake each of the measures.

Procedures

Overview

Participants in this study were asked to take either the adaptive version or non-adaptive version of NCAPS twice. They first answered the measure honestly, then took the measure a second time purposely trying to inflate their scores. A previous study by McFarland and Ryan (2000) found that the order of instructions (e.g., fake first or honest first) did not affect the results. For ease of administration, participants were asked to take the honest condition first. Differences of the personality scores from the honest to fake conditions were compared between the adaptive form and the non-adaptive form. It was hypothesized that participants would have more difficulty purposely inflating their scores on the forced-choice paired-comparison adaptive NCAPS version for reasons previously offered.

Participants

Participants for this study were recruited from introductory psychology courses and online wellness courses from an urban university. Students were offered extra course credit for participation in the study. Students in the introductory psychology courses are typically college students between 19 and 21 years of age. To get a more comprehensive sample, students from the online wellness course were also recruited, because these students are typically non-traditional students from a more age-diverse background. A total of 158 students participated. The ages of the participants ranged from 17 to 53, with 70 percent of the participants in the 17–21 age range. The gender makeup of the participants was 73 percent female and 27 percent male. The percentages for ethnic makeup of the participants were: 62% Caucasian, 32% Black, 3.5% Other, 2% Asian, and the remaining 0.5% were either non-respondents or Hispanic.

Measures and Materials

All measures for the study were completed by the students online via a secured internet connection. At the time of recruitment, students gave the researchers their names and university e-mail address. This information was entered into the database to verify credentials at login. Students were given an internet address for the study. In order to access the study measures, they were required to enter the information they previously supplied to the researchers. Once login credentials were verified, they were presented with an informed consent and instructions for participating in the study.

Participants were assigned to take either a traditional single-statement Likert-scale version or the paired-comparison adaptive format NCAPS. The traditional format consisted of 172 personality statements with 5-point Likert-scale responses (i.e., Strongly Agree to Strongly Disagree). These personality statements represented 10 personality dimensions. The order of the statements was arranged so that items for each of the constructs were interspersed and not presented together. Scale reliabilities ranged from .68 to .84 with most .70 and higher. Please see Table A-1 in the Appendix for the number of items per construct and scale reliabilities.

The adaptive version of NCAPS is a paired-comparison forced-choice measure that uses item response theory (IRT) methodology to improve score accuracy by selecting items for presentation that are tailored to a respondent's ability or personality level. Participants were presented with a total of 120 unidimensional paired-comparison statements. Twelve pairs of statements were presented for each of the 10 personality constructs being measured. The constructs were interspersed randomly during the test so that the item pairs for each construct were not presented together. The first pair of statements for a construct represented mid-level trait scores. Once an item was chosen, the next pair of statements for that construct had trait levels that bracket the examinee's score on the last pair. Item presentation continued in this manner until 10 pairs per construct were presented. This is just a synopsized explanation of the mechanisms behind NCAPS administration and scoring. For a more detailed description of the adaptive theory and functioning of NCAPS please refer to Stark, Chernyshenko, and Drasgow (2006), Stark and Drasgow (2002), and Underhill (2006).

Participants were also asked to take a cognitive ability test called the Wonderlic Quick Test (WPT-Q) which is an 8-minute internet version of the Wonderlic Personnel Test (WPT). The WPT-Q was developed by Wonderlic, Inc. to reliably measure cognitive ability in an unsupervised internet environment. Wonderlic, Inc. has reported the internal reliability of the WPT-Q as $\alpha = .81$ and a corrected correlation with the full length WPT as $r = .93$ (Wonderlic, 2004).

Design and Study Procedures

As students logged into the experiment, they were alternatively assigned to one of two conditions or formats (e.g., traditional format or adaptive format). The procedures for both format groups were the same; see Table 2 for the actual instruction text. Participants were first instructed to take the personality measure honestly. After completion of the first measure, they were then given instructions to take the same measure again as if they were applying for a job and wanted to make the best impression possible. They were instructed to "fake good" their results. At the completion of the personality measure in the second condition, the participants were provided a hyperlink to the secure site on which to take the WPT-Q. Results of the WPT-Q were sent to the researcher. The total experiment time ranged from 45 minutes to one and a half hours.

Table 2
Instructions to Participants

	Honest Instructions	Faking Instructions
Traditional Format	This survey contains statements describing opinions, feelings, or behaviors. For this first administration we are asking you to read each statement carefully and answer HONESTLY. Using the scale provided, indicate how accurately each statement describes you as you generally are now, not as you wish to be. Please respond as accurately and honestly as possible. There are no "correct" or "incorrect" answers. We have also found that it is best to work at a fairly rapid pace, so don't spend too much time on one question.	In this next and last administration of NCAPS we are asking you to read each statement and answer as if you were applying for a job. Please don't answer honestly. Deliberately answer in a way that would make you look more favorable in order to make the best impression possible.
Adaptive Format	This survey contains pairs of statements. Each of these statements describes an opinion, feeling, or behavior. For this first administration, carefully read each pair and decide which statement most accurately describes you as you generally are now, not as you wish to be. Respond as accurately and HONESTLY as possible. There are no "correct" or "incorrect" answers. We have also found that it is best to work at a fairly rapid pace, so don't spend too much time on each pair.	In this next and last administration of NCAPS we are asking you to read each pair of statements and answer as if you were applying for a job. Please don't answer honestly. Select the statement that would make you look more favorable in order to make the best impression possible.

Data Scoring

The traditional version was scored by the same method used for scoring the data from the previous NCAPS pilot tests (Ferstl et al., 2003; Underhill, 2004). Items on the traditional format of NCAPS came from the entire NCAPS item pool whose items represent varying levels of traits along a 2 to 8 scale. Computations were made to standardize responses based on each item's trait level and a person's response to that item. For example, someone's response "strongly agree" to an item that is rated a trait value of 3 (e.g., "I try to do my best at some things") is not equivalent to his or her response "strongly agree" to an item representing a trait value of 7 (e.g., "I excel at virtually everything I try"). Once standardized responses were calculated for each participant, items for each construct were summed to get an overall trait score for each personality dimension. The adaptive NCAPS program scores and revises participants' individual personality construct scores as they respond to each item pair using the adaptive IRT methodology previously mentioned.

Data Integrity

The integrity of the data was examined by looking at completeness of responses as well as outlier detection. Four participants had incomplete data on the adaptive NCAPS. These four participants were removed from analyses. Personality scores from each instruction group within each format group were converted to z-scores. First, scores in the honest instruction condition were examined for z-scores greater than 3. Second, scores in the faking instruction were examined for z-scores of 2.5 or higher. Five participants were removed from analysis because of consistently high z-scores which indicated abnormal responding in relation to the group responses in the faking or honest instruction conditions.

Group Differences

There were no significant demographic differences between the traditional and adaptive groups. Ages of the participants in the adaptive group ranged from 17 to 53, with a mean of 21. In the traditional group the ages ranged from 18 to 36 with a mean of 20. Males and females were evenly distributed between adaptive (males = 21, females = 55) and traditional (males = 19, females = 53) groups.

Results

In each format (i.e., adaptive or traditional), participants were asked to take the measure honestly then asked to fake it or try to make the best impression possible (see Table 2 for the instructions). Higher personality trait scores in the faking condition than in the honest condition would indicate intentional response distortion. Differences between honest and faking scores were analyzed separately for the adaptive and traditional measures. Paired-comparison *t*-tests were conducted for scores on each of the ten personality constructs. The experimentwise alpha was adjusted to account for

any capitalization on chance which may occur when multiple comparisons are made. The Bonferroni correction of dividing the experimentwise alpha of .05 by the number of comparisons made was done for each format group to determine the level of significance to be met for each *t*-test (Pedhazur, 1997, p. 385).

There were no significant mean differences between honest and faking scores on any of the 10 personality traits measured by the adaptive format NCAPS.[3] There were however, significant mean differences between honest and faking scores on all 10 traits measured by the traditional format NCAPS. These striking results show that participants were not able to intentionally distort their personality scores when taking the adaptive format NCAPS. Participants were able to dramatically and significantly increase their personality scores on the traditional (Likert-scale) format (see Table 3).

Table 3
Mean scores and standardized mean differences

	Adaptive NCAPS n = 75				Traditional NCAPS n = 71			
Personality Trait	Honest	Fake	Diff (F-H)	Effect Size	Honest	Fake	Diff (F-H)	Effect Size
Adaptability Flexibility	6.24	6.23	-.005	-0.011	54.86	71.16	16.30*	5.930
Attention to Detail	6.52	6.39	-.128	-0.152	54.96	68.43	13.46*	4.424
Achievement Motivation	6.18	6.19	.013	0.011	53.00	63.94	10.93*	4.092
Dependability	6.43	6.49	.059	0.064	50.16	57.61	17.45*	2.564
Dutifulness Integrity	6.43	6.39	-.031	-0.045	67.77	81.83	14.05*	4.697
Social Orientation	6.16	6.15	-.008	-0.011	78.24	98.71	20.47*	5.686
Self-reliance	5.56	5.47	-.091	-0.104	51.28	56.16	4.87*	1.876
Stress Tolerance	6.14	6.26	.122	0.123	51.66	70.78	19.11*	6.601
Vigilance	6.19	6.42	.226	0.247	44.52	56.31	11.78*	4.458
Willingness to Learn	6.47	6.40	-.071	-0.079	66.86	80.52	13.66*	4.632

* = significant at the .005 level. (Computed .05/10)

The standardized mean difference effect sizes were computed for each trait and condition. The adaptive format NCAPS produced small effect sizes. There were several traits for which the faking condition produced lower mean scores than the honest condition. The traditional format NCAPS produced large effect sizes for many of the traits demonstrating that faking on the traditional Likert-scale personality items produced significant increases in personality scores. (The group sizes and standard deviations used in the formulas can be found in Tables A-5 and A-6 in the Appendix.)

[3] At the time this document was originally written, there were not good Navy estimates for these traits. In the summer of 2008, with over 22,000 active duty Navy participants, we have good comparative data. Generally, the college student trait means were slightly higher on all 10 traits (average of 0.35) with trait increases ranging from 0.13 to 0.53 points higher (on the 2.0-8.0 scale). The most important point is that there is adequate score scale range for the college student traits scores to both increase and decrease under the faking instructions.

Cognitive Ability and Reading Ability

Multiple linear regressions were performed to examine the role of cognitive and reading ability in a person's ability to fake the traditional Likert-scale version of NCAPS. A multiple linear regression was done for each trait's fakability scores, defined as the faked score minus the honest score for a particular trait. The Wonderlic Cognitive ability score and the ACT reading comprehension score (see Tables A-4 and A-5 in the Appendix) were entered stepwise (*p* to include.05; *p* to delete .10) into a regression model with trait fakability as the dependent variable. Regression models for fakability of nine traits were significant, with cognitive ability being a significant predictor of faking for eight out of the nine personality traits. Reading ability was the only single significant predictor of a person's ability to fake *willingness to learn*. The model for predicting the ability to fake the trait *self reliance* was not significant.

Cognitive ability and reading ability were both significant predictors of faking on *achievement motivation*. Together these two predictors explained 50 percent of the variance in fakability of *achievement motivation*. Cognitive ability alone predicted 56 percent of variance in faking scores of *dependability*. Among the other six traits (excluding *willingness to learn* and *self reliance*), cognitive ability significantly explained between 16 percent and 39 percent of the variance in faking (see Table 4).

Table 4
Regression model statistics

Trait	Mean Fakability	Std Dev	Predictor	r^2	Adj r^2	F change	Std Error	Sig
Adaptability Flexibility	17.61	10.35	Cog	.271	.233	7.07	9.06	.015**
Attention to Detail	14.23	8.49	Cog	.382	.349	11.74	6.85	.003**
Achievement Motivation	11.38	6.30	Cog	.421	.390	13.80	4.92	.001**
Achievement Motivation			Read	.551	.501	5.23	4.44	.03**
Dependability	16.17	9.13	Cog	.585	.563	26.80	6.03	.000**
Dutifulness Integrity	13.53	6.71	Cog	.238	.198	5.95	6.00	.025**
Self Reliance*	5.11	7.36	Cog	.180	.137	4.17	6.83	.055
Social Orientation	22.54	16.31	Cog	.242	.202	6.06	14.57	.024**
Stress Tolerance	21.00	11.79	Cog	.209	.137	5.01	10.76	.037**
Vigilance	12.55	6.81	Cog	.209	.167	5.02	6.21	.037**
Willingness to Learn	15.21	8.24	Read	.204	.162	4.85	7.55	.04**

* The stepwise criteria for this model were increased to (*p* to include .10 and *p* to delete .15).
** Significant values.

All regression models were examined for normality and residual outliers and were not found to violate any assumptions. There was a linear relationship between the dependent and independent variables for all models. This was confirmed by plotting the residuals against unstandardized predicted values. Residuals greater than an absolute value of 2.5 were evaluated by the Cooks Distance statistic and by plotting changes in predicted values when cases were deleted from the model. No identified residual outlier had a significant impact on the predicted value in the regression models.

Discussion

Average scores on the NCAPS personality traits suggest the adaptive NCAPS forced-choice format to be far more resistant to faking than the traditional NCAPS Likert-scale format. Participants taking the traditional Likert-scale NCAPS were able to intentionally distort all 10 trait scores, whereas those taking the adaptive paired-comparison NCAPS were not able to significantly distort a single trait score. This study serves to corroborate results by Martin et al. (2002) and Jackson et al. (2000) in which both studies showed the forced-choice format to be more difficult to fake than a Likert scale.

Further analyses regarding individual differences in faking ability showed that all participants had a difficult time faking results on the adaptive paired-comparison NCAPS, even those with high *achievement motivation* or low *integrity* scores. On the other hand, on the traditional Likert-scale version of NCAPS, participants higher in cognitive ability and reading ability were able to produce high fakability scores. Combined, these results support the notion that, regardless of the intelligence or reading levels associated with those taking the adaptive NCAPS; it will be difficult to fake the adaptive paired-comparison format. Therefore, the adaptive paired-comparison NCAPS is more likely to provide results closer to the true trait level scores of the individual rather than falsely inflated scores intended to help an individual get hired or obtain a specific job.

A few potential drawbacks to the study include the generalizability of the results to the Navy population. Since the study was conducted with university students, the majority of whom were female, an argument could be made that the students are a specialized sample and therefore the results will not generalize to the Navy population. However, the results also showed that the higher cognitive and reading ability scores, commonly associated with a sample of college students, did not have any correlation with the fakability of the adaptive paired-comparison NCAPS. Therefore, the likelihood is high that the results will generalize beyond the college student sample to the Navy population. Even so, future research studies should address this concern in a sample of Navy recruits.

References

Bartram, D. (1993). Emerging trends in computer-assisted assessment. In H. Schuler, J. L. Farr, & M. Smith (Eds.), *Personnel selection and assessment: Individual and organizational perspectives* (pp. 267-288). Hillsdale, NJ: Lawrence Erlbaum Assoc.

Borman, W. C., Hedge, J. W., Ferstl, K. L., Kaufman, J. D., Farmer, W. L., & Bearden, R. M. (2003). Current directions and issues in personnel selection and classification. In J. J. Martocchio & G. R. Gerris (Eds.) *Research in personnel and human resource management* (vol. 22). Amsterdam: Elsevier.

Ferstl, K. L., Schneider, R. J., Hedge, J. W., Houston, J. S., Borman, W. C., & Farmer, W. L. (2003). *Following the roadmap: Evaluating potential predictors for Navy selection and classification* (Technical Report No. 421). Minneapolis, MN: Personnel Decisions Research Institutes.

Gottfredson, L. S. (Ed.) (1986). The g factor in employment (Special issue). *Journal of Vocational Behavior*, **29** (3).

Houston, J. S., Borman, W. C., Farmer, W. L., & Bearden, R. M. (Eds.). (2005). *Development of the Enlisted Computer Adaptive Personality Scales (ENCAPS), Renamed Navy Computer Adaptive Personality Scales (NCAPS)* (Tech. Report No. 503). Minneapolis, MN: Personnel Decisions Research Institutes, Inc.

Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance, 11* (2/3), 209-244.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72-98.

Jackson, D. C., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance, 13* (4), 371-388.

Lautenshlager, G. J. (1986). Within-subject measures for the assessment of individual differences in faking. *Educational and Psychological Measurement, 46*, 309-317.

McFarland, L. A. & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85* (5), 812-821.

Martin, B. A., Bowen, C. C., & Hunt, S. T. (2002). How effective are people at faking on personality questionnaires? *Personality and Individual Differences, 32*, 247-256.

McCloy, R. A., Campbell, J. P., & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. *Journal of Applied Psychology, 78*, 493-505.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology*, **43**, 335-354.

Mersman, J. L. & Shultz, K. S. (1998). Individual differences in the ability to fake on personality measures. *Personality and Individual Differences, 24* (2), 217-227.

Mueller-Hanson, R., Heggestad, E. D., & Thornton, III, G. C. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology, 88* (2), 348-355.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction (3rd ed.)*. New York: Harcourt Brace College Publishers.

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than g. *Journal of Applied Psychology, 79* (4), 518-524.

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83* (4), 634-644.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124* (2), 262-274.

Stark, S., Chernyshenko, S., & Drasgow, F. (July 2006). Examination of the computerized adaptive NCAPS program (Drasgow Consulting Group Report to the Navy). Millington, Tennessee: Navy Personnel Research, Studies, and Technology.

Stark, S. & Drasgow, F. (2002). An EM approach to parameter estimation for the Zinnes and Griggs paired-comparison IRT model. *Applied Psychological Measurement, 26* (2), 208-227.

Underhill, C. (2004). *Comparison of two methods of item-pair presentation in a forced-choice computer adaptive personality instrument.* Unpublished doctoral dissertation, The University of Memphis, Memphis, TN.

Underhill, C. (2006). *Investigation of item-pair presentation and construct validity of the Navy Computer Adaptive Personality Scales (NCAPS)*. (NPRST-TN-06-9). Millington, TN: Navy Personnel Research, Studies, and Technology.

Wainer, H. (Ed.) (2000). *Computer-Adaptive Testing: A Primer.* Mahwah, NJ: Lawrence Erlbaum.

Wainer, H., & Mislevy, R. J. (2000). Item response theory, item calibration and proficiency estimation. In H. Wainer et al., *Computerized adaptive testing: A primer (*2nd ed., pp. 61-100). Mahwah, NJ: Lawrence Erlbaum Assoc.

Wainer, H., Dorans, N. J., Green, B. F., Mislevy, R. J., Steinberg, L., & Thissen, D. (2000). Future challenges. In H. Wainer et al., *Computerized adaptive testing: A primer (*2nd ed., pp. 231- 269). Mahwah, NJ: Lawrence Erlbaum Assoc.

Wonderlic, Inc. (2004). *The Wonderlic QuickTest series of tests successfully predicts scores on the Wonderlic Personnel Test (WPT)* (Research Note). Libertyville, IL: Wonderlic, Inc.

Zickar, M. J., Gibby, R. E., & Robie, C. (2004). Uncovering faking samples in applicant, incumbent, and experimental data sets: An application of mixed model item response theory. *Organizational Research Methods, 7* (2), 168-190.

Appendix

Table A-1
Scale reliabilities for the Traditional Format of NCAPS n = 77

Construct	Scale Mean	STD	# of items	Alpha
Achievement Motivation	53.31	6.78	15	.784
Stress Tolerance	52.29	8.63	18	.749
Social Orientation	78.53	11.84	23	.848
Adaptability Flexibility	55.44	6.83	18	.760
Attention to Detail	55.49	8.59	16	.837
Dependability	50.72	8.26	15	.798
Dutifulness and Integrity	68.27	8.23	19	.791
Self-reliance	51.33	6.75	16	.762
Willingness to Learn	67.04	7.09	19	.689
Vigilance	44.77	6.38	13	.787

Table A-2
Adaptive

	M	STD	Variance	Min	Max
Adaptability Flexibility					
Honest	6.24	.732	.535	3.60	7.21
Faking	6.23	.793	.629	3.78	7.42
Attention to Detail					
Honest	6.54	.652	.424	3.96	7.57
Faking	6.39	.809	.655	3.88	7.46
Achievement Motivation					
Honest	6.19	.711	.506	3.71	7.36
Faking	6.19	.808	.653	3.54	7.33
Dependability					
Honest	6.44	.896	.803	3.84	7.45
Faking	6.49	.860	.740	4.30	7.46
Dutifulness, Integrity					
Honest	6.43	.812	.661	3.98	7.35
Faking	6.39	.737	.543	4.60	7.41
Social Orientation					
Honest	6.18	.831	.691	4.04	7.24
Faking	6.15	.792	.627	4.05	7.29
Self-reliance					
Honest	5.55	.707	.500	4.11	7.17
Faking	5.47	.801	.641	3.12	7.52
Stress Tolerance					
Honest	6.13	.914	.835	3.67	7.45
Faking	6.26	1.00	1.00	3.36	7.37
Vigilance					
Honest	6.21	.841	.706	3.57	7.48
Faking	6.42	.895	.801	4.16	7.56
Willingness to Learn					
Honest	6.5	.734	.539	3.80	7.36
Faking	6.41	.834	.697	2.67	7.42

Honest n = 77 Faking n = 75

Table A-3
Traditional

	M	STD	Variance	Min	Max
Adaptability Flexibility					
Honest	54.96	6.28	39.51	40.50	69.43
Faking	71.16	8.85	78.36	34.53	83.61
Attention to Detail					
Honest	55.14	8.43	71.09	36.04	74.98
Faking	68.43	10.12	102.48	37.92	76.58
Achievement Motivation					
Honest	53.13	6.85	46.96	36.12	67.54
Faking	63.94	7.45	55.61	39.23	72.04
Dependability					
Honest	50.24	8.06	65.09	34.60	68.06
Faking	67.61	8.83	77.96	40.32	74.41
Dutifulness, Integrity					
Honest	67.96	8.08	65.37	47.61	84.02
Faking	81.83	9.85	97.03	47.83	92.66
Social Orientation					
Honest	78.64	12.10	146.55	44.12	107.18
Faking	98.71	13.83	191.22	51.99	114.88
Self-reliance					
Honest	51.27	6.87	47.19	35.5	67.44
Faking	56.16	6.66	44.43	43.02	71.68
Stress Tolerance					
Honest	51.87	8.44	71.30	29.70	72.51
Faking	70.78	8.34	69.60	47.33	86.32
Vigilance					
Honest	44.46	6.39	40.83	26.74	58.07
Faking	56.32	7.61	57.90	30.25	63.02
Willingness to Learn					
Honest	66.97	6.78	46.06	51.99	83.17
Faking	80.52	10.64	113.39	45.22	91.04
Honest n = 72 Faking n = 71					

Table A-4
Wonderlic

	Mean	STD	Range
Adaptive n = 48	21.95	4.51	13–30
Traditional n = 50	23.4	4.58	14–31

Table A-5
ACT Reading Score

	Mean	STD	Range
Adaptive n = 32	20.09	4.36	14-30
Traditional n = 33	23.06	5.44	13-35

Table A-6
ACT Comprehensive Score

	Mean	STD	Range
Adaptive n = 32	19.78	3.11	15–26
Traditional n = 32	22.68	4.26	16–33

Distribution

AIR UNIVERSITY LIBRARY
ARMY RESEARCH INSTITUTE LIBRARY
ARMY WAR COLLEGE LIBRARY
CENTER FOR NAVAL ANALYSES LIBRARY
DEFENSE TECHNICAL INFORMATION CENTER
HUMAN RESOURCES DIRECTORATE TECHNICAL LIBRARY
JOINT FORCES STAFF COLLEGE LIBRARY
MARINE CORPS UNIVERSITY LIBRARIES
NATIONAL DEFENSE UNIVERSITY LIBRARY
NAVAL HEALTH RESEARCH CENTER WILKINS BIOMEDICAL LIBRARY
NAVAL POSTGRADUATE SCHOOL DUDLEY KNOX LIBRARY
NAVAL RESEARCH LABORATORY RUTH HOOKER RESEARCH LIBRARY
NAVAL WAR COLLEGE LIBRARY
NAVY PERSONNEL RESEARCH, STUDIES, AND TECHNOLOGY SPISHOCK LIBRARY (3)
PENTAGON LIBRARY
USAF ACADEMY LIBRARY
US COAST GUARD ACADEMY LIBRARY
US MERCHANT MARINE ACADEMY BLAND LIBRARY
US MILITARY ACADEMY AT WEST POINT LIBRARY
US NAVAL ACADEMY NIMITZ LIBRARY