

Big Five Factors of Personality and Replicated Predictions of Behavior

Sampo V. Paunonen
University of Western Ontario

Measures of the Big Five factors of personality were used to predict a variety of criterion variables thought to represent behaviors of some social and cultural significance (e.g., alcohol consumption, grade point average). Analyses focused on replicated predictions across 2 independent samples of participants ($N_s = 276$ and 142) with 3 different measures of the Big Five (the NEO Five-Factor Inventory, the Revised NEO Personality Inventory, and the Five-Factor Nonverbal Personality Questionnaire, the latter an experimental nonverbal personality inventory). The results indicated substantial consistency in behavior predictions across the different Big Five assessments. The data are interpreted as supporting both the construct validity of the personality measures used and the role of the Big Five factors as determinants of certain complex behaviors.

The five-factor model of personality structure has generated substantial interest among personality researchers. The belief is that personality-based variations in behavior are largely interpretable in terms of the Big Five factors of Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience (see McCrae & John, 1992). This view of personality structure has consequently led to the development of several questionnaires, inventories, and adjective rating scales designed to measure those broad dimensions of behavior (e.g., see De Raad & Perugini, 2002).

My primary purpose in this study is to evaluate the generalizability of a few Big Five personality factor inventories as predictors of a common set of criteria, criteria representing behaviors of some social significance. By assessing people using multiple inventories purporting to measure the Big Five personality factors, one can determine whether the different measures of the same personality constructs show any consistency in their prediction of the same criterion variables. If the Big Five inventories are all designed to measure equivalent dimensions of personality, then they should show a nontrivial amount of agreement in the variables they are able to predict.

I used three measures of the Big Five in this study. Two of these were the well-known NEO Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1992) and Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) questionnaires. The third measure is a new, structured, nonverbal measure of the five personality factors called the Five-Factor Nonverbal Personality Questionnaire (FF-NPQ; Paunonen, Ashton, & Jackson, 2001). The construction of the FF-NPQ has been described in detail elsewhere (Paunonen & Ashton, 2002; Paunonen et al., 2001), so I summarize it only briefly.

This research was supported by Social Sciences and Humanities Research Council of Canada Research Grant 410-98-1555. I thank Paul MacDonald and Andrew Holownych for their assistance in the data collection and Ray Wolfe and Paul Olczak for their comments on a draft of this article.

Correspondence concerning this article should be addressed to Sampo V. Paunonen, Department of Psychology, University of Western Ontario, London, Ontario N6A 5C2, Canada.

The FF-NPQ

The FF-NPQ is a 60-item five-factor measure based on the Nonverbal Personality Questionnaire (NPQ), which is a longer 136-item measure of 16 Murray needs (see Paunonen & Ashton, 1998; Paunonen, Jackson, & Keinonen, 1990; Paunonen, Jackson, Trzebinski, & Forsterling, 1992; Paunonen et al., 1996; Paunonen, Zeidner, Engvik, Oosterveld, & Maliphant, 2000). Most items in the FF-NPQ were selected from the NPQ (a few new items were created) using item analysis and factoring procedures to identify the best Big Five exemplars in the nonverbal item pool (see Paunonen & Ashton, 2002).

Each NPQ and FF-NPQ item represents a line drawing of a central character performing a trait- or factor-relevant behavior in a specific situation. Respondents are asked to consider each item in a questionnaire and to decide, using a 7-point rating scale, the likelihood that they would engage in the type of behavior depicted in the illustration. An example FF-NPQ item (representing thrill-seeking behavior) is shown on the Instructions and Rating Form page reproduced in Figure 1. Other examples of the nonverbal items can be found in the articles by Paunonen and Ashton (2002), Paunonen et al. (1990, 1996, 2001), and Paunonen and Jackson (1979).

As already stated, the items of the FF-NPQ were selected empirically from the longer NPQ to measure the Big Five factors of personality. The particular conceptualization of personality structure chosen for this task was the same five-factor model underlying the construction of the NEO-FFI and the NEO-PI-R. In fact, the domain scales on the NEO-FFI were used as one basis for selecting FF-NPQ items from the nonverbal item pool—NPQ items that correlated well with the particular NEO-FFI scales were assigned to measure those factors in the FF-NPQ (see Paunonen et al., 2001). An implication of this test construction procedure is that one of the FF-NPQ scales measures Openness to Experience, instead of a dimension that others have labeled as Culture or Intellect (see Digman, 1990).

In preliminary studies with the FF-NPQ (Paunonen et al., 2001), the five nonverbal factor scales showed good levels of internal consistency reliability and convergent validity. For example, in a sample of 304 Canadian students, the mean internal consistency of

FF NPQ

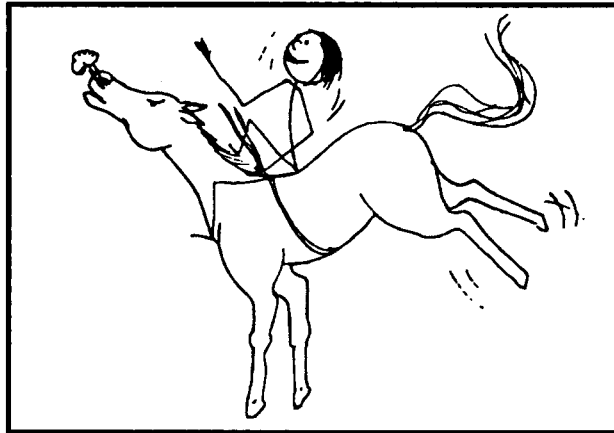
Instructions and Rating Form

Attached is a **Picture Booklet** containing a series of illustrations depicting a *central figure* (the one with the *hair* drawn in) performing specific behaviors in certain situations.

Please look at each illustration and rate the *likelihood that you would engage in the type of behavior* shown.

Using the **Rating Form** on the other side of this page, record your responses by selecting an appropriate number from the 7-point rating scale. Consider the example below:

- 7 - **extremely likely** that I would perform this type of behavior
- 6 - **very likely** that I would perform this type of behavior
- 5 - **moderately likely** that I would perform this type of behavior
- 4 - **neither likely nor unlikely** that I would perform this type of behavior
- 3 - **moderately unlikely** that I would perform this type of behavior
- 2 - **very unlikely** that I would perform this type of behavior
- 1 - **extremely unlikely** that I would perform this type of behavior



Rating. 6

In this example the person has responded that it would be *very likely* that he/she would engage in the kind of activity in which the central figure is engaging. Your own response might have been different.

In a similar manner, consider each illustration in the **Picture Booklet** and estimate the likelihood that you would engage in the type of behavior depicted by the central figure.

Please respond to every picture and record your response on the 7-point rating scales on the back of this page. Do not mark the **Picture Booklet** or any other materials.

Turn Page Over . . .

the scales was .80, and the mean convergence with corresponding NEO–FFI scales was .52. Self-ratings and peer ratings of university residence roommates collected on the FF–NPQ scales also correlated positively, with a mean value of .41. Moreover, both the FF–NPQ and the NEO–FFI were found to show a strikingly similar pattern of relations in predicting different criterion variables. For example, in a set of 14 behavior criteria, both Big Five questionnaires best predicted the same criterion (self-reported alcohol consumption of students), and both had the same mean multiple correlation of scales with criteria (both $R = .25$, each averaged across the 14 criterion variables).

The FF–NPQ, being a nonverbal personality inventory, is of obvious interest to cross-cultural researchers. Because no translation of the items is required (the instructions shown in Figure 1 do, of course, require translation), the questionnaire can be easily administered to people in many different language groups as well as to people who have difficulty reading or understanding any language. Further evaluations of the FF–NPQ items (Paunonen et al., 2001) have shown those Big Five scales to have good levels of internal consistency reliability and reasonable degrees of convergence with verbal indicators of the Big Five in several cross-cultural samples (see also Paunonen & Ashton, 2002).

The Big Five as Predictors

Construct valid measures of personality variables should predict human behaviors, assuming those behaviors have personality determinants. This is especially true of Big Five inventories because those factors are presumed to account for most of the personality-based variation in behavior. But some predictions are more remarkable than are others. For example, personality test scores that correlate with respondent self-ratings using other methods of measuring the same trait (e.g., comparing respondents' FF–NPQ scale scores with their NEO–FFI scale scores, as reported in the previous section) is a mandatory requirement of good psychometric assessment and, as such, is unremarkable. One might also consider unremarkable test scores that correlate with ratings of respondents on the trait by well-acquainted peers, using the same or different inventories (e.g., comparing respondents' self-ratings with roommate ratings on the FF–NPQ scales). What is remarkable, however, is a personality measure that predicts complex human behaviors, behaviors that are not so obviously and simply parallel indicators of the assessed construct.

What are such human behaviors? Examples are alcohol consumption, grade point average, self-perceived level of physical attractiveness, dating behavior, tobacco consumption, participation in sports, and so on. Such individual-differences variables are complex in the sense that they probably have multiple determinants, including genetic factors and environmental influences. But if those determinants include personality variables, then a good omnibus measure of personality should predict those behaviors. Note that such behavior criteria are important variables in the sense that they are of some social and cultural significance and have implications for personal well-being. If specific personality measures were able to predict some of those complex behaviors, then certainly that would be remarkable.

Not a lot of data exist demonstrating the utility of personality trait or factor measures in predicting complex criteria of the type described above (Funder, 2001). A notable exception has been in

the area of industrial and organizational psychology. Dozens of studies, summarized in several meta-analytic reviews, have shown that Conscientiousness is a good predictor of overall job performance in many occupations (e.g., see Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991). Measures related to Conscientiousness have even been found to predict narrower criteria, such as job effort and irresponsible work behavior (Hough, 1992). In the area of educational psychology, Big Five factor measures sometimes have been used successfully to predict academic achievement. Goff and Ackerman (1992) reported that the Conscientiousness factor was positively related to undergraduate students' grade point averages in their study, whereas Extraversion was negatively related. Paunonen and Ashton (2001b) showed that Conscientiousness but not Openness to Experience predicted the grades of undergraduates enrolled in a course on personality theory. Using a different criterion of academic performance—classroom participation grades in a large sample of master's of business administration students—Rothstein, Paunonen, Rush, and King (1994) found significant positive correlations for the Extraversion and Openness to Experience factors but no correlation for Conscientiousness.

Although there have been some demonstrations of the utility of Big Five factor measures in predicting work performance and academic achievement, the better levels of predictive validity usually have been in the neighborhood of only .20. In some earlier work, I presented logical arguments and empirical data showing that the facets of the Big Five factors (i.e., the narrow personality traits that constitute each personality factor) can have even more utility than the factors themselves in predicting such complex, real-life criteria (Paunonen, 1993, 1998; Paunonen & Ashton, 2001a, 2001b; Paunonen & Nicol, 2001; Paunonen, Rothstein, & Jackson, 1999). For example, in the study by Paunonen and Ashton (2001b), a broad measure of Conscientiousness predicted grade point average with a value of .20, but a narrower measure of achievement motivation (a facet of Conscientiousness) predicted that criterion with a value of .26. Even more telling, in the Rothstein et al. (1994) study, Extraversion predicted classroom participation grades with a correlation of .19, whereas a measure of exhibitionism (a facet of Extraversion) predicted the same criterion with a correlation of .33.

Arithmetically combining several narrow trait or facet measures to derive a broad factor measure can have undesirable consequences. Some of the traits might be predictive of a criterion of interest, and others might not. When the predictive and nonpredictive facets are aggregated in the pursuit of their common variance, the trait-specific but criterion-valid variance that exists in the former can be canceled by the trait-specific but nonpredictive variance in the latter. It is for this reason that I have argued that personality researchers and assessors should avoid the temptation to use short global measures of the Big Five personality factors instead of more specific Big Five facet or trait measures. The former assessment strategy might seem expeditious, because four or five factor measures are generally quicker to administer and easier to interpret than are many more component facet measures. But my studies have indicated very real costs in adopting that expedient to personality measurement—less accuracy in behavior prediction and a poorer understanding of behavior determinants (see also Ashton, 1998; Ashton, Jackson, Paunonen, Helmes, & Rothstein, 1995; Mershon & Gorsuch, 1988).

Purpose of This Study

I have two primary goals in this study. First, despite the argument I advanced above that there are predictive benefits to using Big Five facets of personality over Big Five factors as predictors of complex criteria, it is nevertheless important to demonstrate that the latter also have some utility in that regard. Furthermore, as mentioned earlier, different (construct valid) measures of the same factors should show some agreement in the variables they are able to predict. Because I administered three different measures of the Big Five in this study and because I assessed many socially important criterion variables, I was able to evaluate these requisites of measurement validity in this research. Although some of the studies I have already cited have demonstrated predictive utility to certain Big Five factor measures, no study has yet applied such a standard of replication as I describe below to the prediction of such a large set of complex behaviors.

My second goal in this study, related to the first, is to provide further evidence of convergent validity for the new nonverbal Big Five questionnaire, the FF-NPQ. New measures of personality are required to show some degree of correlation with established measures of the same traits. This study provides me with an opportunity to evaluate this aspect of the FF-NPQ's construct validity by comparing the nonverbal Big Five scales against corresponding verbal Big Five scales.

Method

Participants

I administered measures of personality factors and behavior criteria to two different samples of university students. The first sample consisted of 276 introductory psychology students (152 men and 124 women), and the second sample consisted of 142 introductory psychology students (56 men and 86 women). Although I analyzed the data separately by sample, I discuss them together in terms of the consistencies of particular Big Five factors in predicting specific criterion variables.

Materials

The FF-NPQ. Both samples of participants completed the FF-NPQ, which has already been described in the introduction. This questionnaire is a 60-item measure of the Big Five personality factors, with 12 nonverbal items on each factor scale. The items consist of pictorial depictions of personality-relevant behaviors being performed in specific situations. Respondents rate the likelihood that they would engage in the types of behaviors illustrated using 7-point rating scales (see Figure 1).

The NEO-FFI and NEO-PI-R. One sample of participants completed the NEO-FFI, whereas the other sample completed the NEO-PI-R. Although both inventories are measures of the Big Five personality factors, the former is a 60-item questionnaire, and the latter is a 240-item questionnaire. The NEO-FFI was constructed to measure the five factors with 12 items each. The much longer NEO-PI-R consists of 30 eight-item facet scales, 6 of which are summed to derive each of the Big Five factor measures. The items on both inventories are rated by respondents on 5-point scales. Note that the 60 items of the NEO-FFI constitute a subset of the 240 items in the NEO-PI-R.

The Behavior Report Form. All participants completed the Behavior Report Form, which is a measure I designed to assess several complex behaviors of some social significance. I first used it in Paunonen (1993) and Paunonen (1998) and later expanded and revised it (Paunonen &

Ashton, 2001a). The four separate pages of the Behavior Report Form are reproduced in the Appendix. For the present analysis, I transformed a few of the behaviors measured by that form somewhat, for reasons that become apparent below (e.g., for Item 18, the total of all medications taken regularly was computed). I also omitted a few variables because of excessive missing data (e.g., for Item 12, almost no students were members of a fraternity) or highly skewed data (e.g., for Item 20, very few students worked many hours at a part-time job). In addition, some variables were not analyzed because they did not represent behaviors as such (e.g., the rating of eyesight in Item 19) or were redundant with other variables (e.g., in Item 9, the number of cigarettes smoked per day was highly redundant with the smoking behavior rating).

The specific Behavior Report Form variables evaluated in this study and used as behavior criteria number 27 in all. These criteria, with their labels in italics, include (1) self-perceived *attractiveness*, (2) self-perceived *intelligence*, (3) self-perceived *popularity*, (4) self-perceived *femininity*, (5) self-perceived *religiosity*, (6) self-perceived *honesty*, (7) *grade point average*, (8) *dating frequency* as number of dates per month, (9) *dating variety* as number of different people dated over a year, (10) *tobacco consumption* as the 5-point rating of smoking behavior, (11) *alcohol consumption* as the 5-point rating of drinking behavior, (12) preference for a *liberal arts* program of study, (13) 9-point rating of *interest in fraternity* or sorority membership, (14) mean number of all *traffic violations* received per year, (15) *driving fast* as the maximum speed driving a car, (16) money spent *buying lottery tickets* per month, (17) number of *parties attended* per month, (18) having engaged in long-term *dieting behavior*, (19) total number of prescription *medications used* regularly, (20) *preference for contacts* over glasses as prescription lenses, (21) engaged in *part-time work*, (22) ability to *play musical instruments*, (23) *participation in sports*, (24) performing *routine exercise*, (25) *donating blood* occasionally, (26) an *obesity index* defined by weight/height², and (27) *participant sex* or gender. (Sex was also used in this study as a covariate in analyses of the other criteria.) Note that four of the criteria listed above were computed variables, based on combinations of two or more Behavior Report Form items (i.e., Criteria 14, 19, 20, and 26).

Note also that the first six criterion variables listed above are each qualified by the term *self-perceived*. This labeling is to highlight the fact that ratings on those variables do not necessarily reflect objective levels of the nominal attributes (i.e., attractiveness, intelligence, popularity, femininity, religiosity, and honesty). Instead, those ratings are best thought of in this study as representing behaviors related to the self-evaluation of those physical or psychological characteristics that, for a variety of reasons, might not agree well with more objective assessments. To illustrate, self-perceived physical attractiveness could be a complexly determined belief, based on one's interpersonal experiences, personal standard of beauty, social comparison network, self-enhancement tendency, and so forth, including some element of objective beauty. This issue of self-perception versus objective assessment and the validity of the Behavior Report Form variables is raised again in the Discussion section of this article.

Procedure

I administered the Big Five personality factor measures and the behavior criterion measures to the two samples of university undergraduates. The first sample ($n = 276$) completed the FF-NPQ, the NEO-FFI, and the Behavior Report Form. Because of an oversight on the part of a research assistant, data were collected from only 178 of the respondents with the FF-NPQ. Also, 3 students failed to complete the NEO-FFI. The second sample ($n = 142$) completed the FF-NPQ, the NEO-PI-R, and the Behavior Report Form. Seven students in that sample failed to complete the NEO-PI-R.

Results

Big Five Scale Summary Statistics

Table 1 lists some of the statistics obtained from the four personality assessments of participants using the FF-NPQ (two assessments), the NEO-FFI, and the NEO-PI-R. Comparison of the scale means for the three different inventories shows that, although no measure was conspicuously far from the midpoint of its items' rating scale (7-point rating scale for the FF-NPQ items, and 5-point rating scale for the NEO-FFI and NEO-PI-R items), the Neuroticism scale scores were below that midpoint in all four assessments, whereas all other scale scores were above the midpoint. The apparent differences in the standard deviations of the factors' mean scores between the nonverbal measures and the verbal measures most likely reflect the differences in the item rating scales used by the inventories (i.e., 7-point vs. 5-point rating scales, respectively).

The alpha coefficients of internal consistency are shown for each Big Five scale in the last column of Table 1 and indicate good levels of reliability for all measures. Note that the NEO-PI-R scales had higher reliabilities, in general, than did the other scales. This is no doubt due to the length of the former scales, at 48 items each, compared with only 12 items each for the other scales. Observe also that the FF-NPQ reliabilities were slightly higher than were the NEO-FFI reliabilities for the scales of Extraversion, Openness, and Agreeableness but slightly lower for Neuroticism and Conscientiousness.

Table 1
Summary Statistics for Big Five Personality Factor Measures

| Measure | No. items | <i>M</i> | <i>SD</i> | α |
|----------------------------|-----------|----------|-----------|----------|
| FF-NPQ (<i>n</i> = 178) | | | | |
| Neuroticism | 12 | 3.61 | 1.01 | .81 |
| Extraversion | 12 | 4.36 | 1.08 | .87 |
| Openness | 12 | 4.65 | 1.02 | .85 |
| Agreeableness | 12 | 5.21 | 1.04 | .85 |
| Conscientiousness | 12 | 4.52 | 0.75 | .79 |
| FF-NPQ (<i>n</i> = 142) | | | | |
| Neuroticism | 12 | 3.85 | 0.98 | .80 |
| Extraversion | 12 | 4.16 | 0.92 | .81 |
| Openness | 12 | 4.63 | 0.91 | .82 |
| Agreeableness | 12 | 5.15 | 0.91 | .80 |
| Conscientiousness | 12 | 4.74 | 0.76 | .74 |
| NEO-FFI (<i>n</i> = 273) | | | | |
| Neuroticism | 12 | 2.65 | 0.68 | .86 |
| Extraversion | 12 | 3.66 | 0.49 | .79 |
| Openness | 12 | 3.36 | 0.52 | .72 |
| Agreeableness | 12 | 3.65 | 0.50 | .76 |
| Conscientiousness | 12 | 3.45 | 0.60 | .86 |
| NEO-PI-R (<i>n</i> = 135) | | | | |
| Neuroticism | 48 | 2.97 | 0.40 | .89 |
| Extraversion | 48 | 3.50 | 0.36 | .87 |
| Openness | 48 | 3.42 | 0.36 | .86 |
| Agreeableness | 48 | 3.41 | 0.34 | .87 |
| Conscientiousness | 48 | 3.27 | 0.36 | .87 |

Note. Mean scores represent the average 7-point item rating for the FF-NPQ scales and the average 5-point item rating for the NEO-FFI and NEO-PI-R scales. FF-NPQ = Five-Factor Nonverbal Personality Questionnaire; NEO-FFI = NEO Five-Factor Inventory; NEO-PI-R = Revised NEO Personality Inventory.

Big Five Scale Intercorrelations

Table 2 shows the monomethod-heterotrait correlations of the Big Five factors measured using different inventories. The overall level of scale relatedness in each of the four data sets was about the same, as indicated by the size of the first eigenvalue of each correlation matrix (1.82, 1.99, 1.74, and 1.89, respectively). However, some differences can be discerned between the specific correlations among the FF-NPQ scales and those among the two verbal measures. For example, Neuroticism was consistently correlated negatively with the other four factor scales in both the NEO-FFI and the NEO-PI-R inventories (consistent with a desirability response bias interpretation). In the case of the FF-NPQ, however, those correlations were slightly positive, even noticeably positive in the case of Neuroticism's correlation with Conscientiousness (consistent with an acquiescence bias interpretation—with the exception of the Agreeableness scale, the items of the FF-NPQ scales are all positively keyed.) Other differences were that Extraversion and Agreeableness were positively correlated in the verbal scales but were uncorrelated in the nonverbal scales and that Openness and Conscientiousness were correlated in the FF-NPQ inventory but not in the NEO-FFI or NEO-PI-R inventories.

The differences in scale correlations described above suggest that the nonverbal items and the verbal items may be measuring somewhat different aspects of the same Big Five factors. However, despite these differences in heterotrait correlations, the FF-NPQ correlated .55 on average with the corresponding NEO-FFI factor scales in the first sample's data (Neuroticism = .57, Extraversion = .51, Openness = .64, Agreeableness = .48, Conscientiousness = .56; in contrast, the mean absolute value of the heteromethod-heterotrait correlations was appropriately low at only .11). Coincidentally, the FF-NPQ also correlated .55 with the NEO-PI-R scales in the second sample's data (Neuroticism = .51, Extraversion = .57, Openness = .63, Agreeableness = .58, Conscientiousness = .46, with a mean absolute heteromethod-heterotrait correlation of only .17). These convergent validity values between the verbal and nonverbal scales are not much lower than the .59 mean correlation reported between NEO-FFI factor scales and adjective-based Big Five marker variables (Costa & McCrae, 1992, p. 54).

The results of the heterotrait and monotrait comparisons for the present Big Five inventories suggest a convergence of the corresponding factor measures. However, the verbal and nonverbal scales are probably measuring slightly different variants of the same personality dimensions. As has been noted elsewhere (e.g., Paunonen et al., 1992), certain behaviors are more easily represented with verbal items than with nonverbal items. This limits, to some extent, the range of trait content that can be incorporated in nonverbal personality scales of the type found in the FF-NPQ.

Replicated Significant Big Five Predictors

As an overall summary of how well the Big Five inventories could predict the present criteria, I first computed the multiple correlations between the five factor scales in an inventory and each behavior criterion. Averaged across all 27 criteria used in this study, the mean multiple correlations for the two FF-NPQ data sets were .27 and .28 (*n* = 178 and *n* = 142, respectively). The corresponding mean multiple correlations were .27 for the NEO-

Table 2
Scale Intercorrelations for Big Five Personality Factor Measures

| Measure | 1 | 2 | 3 | 4 | 5 |
|----------------------------|---|------|------|------|------|
| FF-NPQ (<i>n</i> = 178) | | | | | |
| 1. Neuroticism | — | .05 | .06 | .11 | .27 |
| 2. Extraversion | — | — | .18 | -.05 | .05 |
| 3. Openness | — | — | — | .37 | .38 |
| 4. Agreeableness | — | — | — | — | .34 |
| 5. Conscientiousness | — | — | — | — | — |
| FF-NPQ (<i>n</i> = 142) | | | | | |
| 1. Neuroticism | — | .16 | .14 | .05 | .38 |
| 2. Extraversion | — | — | .22 | -.08 | .18 |
| 3. Openness | — | — | — | .42 | .48 |
| 4. Agreeableness | — | — | — | — | .29 |
| 5. Conscientiousness | — | — | — | — | — |
| NEO-FFI (<i>n</i> = 273) | | | | | |
| 1. Neuroticism | — | -.33 | -.14 | -.14 | -.18 |
| 2. Extraversion | — | — | .17 | .27 | .19 |
| 3. Openness | — | — | — | .09 | -.02 |
| 4. Agreeableness | — | — | — | — | .25 |
| 5. Conscientiousness | — | — | — | — | — |
| NEO-PI-R (<i>n</i> = 135) | | | | | |
| 1. Neuroticism | — | -.26 | -.14 | -.17 | -.27 |
| 2. Extraversion | — | — | .40 | .33 | .11 |
| 3. Openness | — | — | — | .26 | .11 |
| 4. Agreeableness | — | — | — | — | .12 |
| 5. Conscientiousness | — | — | — | — | — |

Note. FF-NPQ = Five-Factor Nonverbal Personality Questionnaire; NEO-FFI = NEO Five-Factor Inventory; NEO-PI-R = Revised NEO Personality Inventory.

FFI (*n* = 273) and .33 for the much longer NEO-PI-R (*n* = 135). These multiple correlations indicate some degree of criterion predictiveness to the factor measures overall, but they do not indicate the extent to which the three personality inventories were able to replicate each other in the prediction of specific criterion variables across the four data sets.

I chose to identify replicated factor predictors of the present behavior criteria by comparing correlations of each Big Five factor scale with each criterion across the four personality assessments. These were partial correlations, in which participant sex or gender was removed. Simple sex differences on the criterion and predictor variables could contribute to a correlation between those variables that does not exist within the sexes. (Gender was not partialled, of course, when participant sex was the criterion variable.)

For these analyses, I had to set some standard of replication in factor-criterion correlations across the four personality assessments (i.e., the NEO-FFI, NEO-PI-R, and two FF-NPQ assessments). I decided on the following arbitrary decision rule: If the four correlations for a particular factor-criterion combination were all in the same direction and at least three of the four were statistically significant ($p < .05$), then I would conclude that the empirical finding replicated across the four data sets. When this decision rule is used, the Type I error rate is appropriately small for each replication evaluation. The probability of finding four out of four empirical tests significant, with a predictor-criterion association that is actually zero in the population, is .00000625. When one allows for three or four tests to be significant, that probability is still small at .00048125. Furthermore, the experimentwise Type

I error rate for this procedure, considered across all five factor predictors and 27 criteria, is less than .065.

Table 3 lists the replicated factor-criterion correlations found across the four data sets. Overall, 12 of the 27 criteria were predicted by the same factors in each of the four assessments. Those criteria are ordered in the table roughly in terms of the strength of the associations. For example, sex differences on the factor scales were the strongest predictors of participant gender, with the men in the samples being consistently lower in Neuroticism, Agreeableness, and Conscientiousness than were the women. Self-perceptions of intelligence were also consistently and well predicted by scores on the Neuroticism (low), Openness, and Conscientiousness factor scales. The Big Five factor scales in the present data were further able to predict such important variables as alcohol consumption (high Extraversion, low Conscientiousness) and tobacco consumption (low Agreeableness).

Replicated Nonsignificant Big Five Predictors

In evaluating the generalizability of different Big Five factor measures as predictors of the present criteria, it is also relevant to consider the null results. That is, it is important to know the extent to which the factor scales are in agreement in not being able to predict a particular criterion. For this evaluation, I used the following arbitrary decision rule for a replicated null finding: for each factor-criterion comparison, all four correlations nonsignificant ($p > .05$) or, at most, one significant correlation and three nonsignificant.

One way to guarantee replicated nonsignificant results, of course, is to ensure low power in one's statistical tests. Low power, however, was not a major issue in the present analyses. Assuming a small to medium effect size for a factor-criterion correlation of .20 in the population, the probabilities of correctly detecting that association (i.e., power) in the four data samples of this study (with $\alpha = .05$) are about .86 for $n = 178$, .77 for $n = 142$, .95 for $n = 273$, and .75 for $n = 135$ (Cohen, 1988, p. 87). Therefore, the probabilities of erroneously finding a null result (i.e., committing a Type II error) in any single statistical test with those four samples are about .14, .23, .05, and .25, respectively.

Using the lowest estimated level of power for the individual statistical tests in this study (i.e., .75 for $n = 135$), one can compute the probability of making a Type II error when using the conjoint decision rule I set for replicated null results. That is, if one accepts either three or four nonsignificant findings as a replicated null effect, what is the probability that one would erroneously conclude that the factor-criterion correlation is zero when in fact it is .20 in the population? The probability of obtaining exactly four out of four nonsignificant statistical tests under the present conditions is less than .0039. The probability of three or four nonsignificant results is still acceptably low at less than .0507. These values represent liberal estimates of the present Type II error rate because they are based on only the data sample yielding the least powerful statistical tests.

Applying my standard of (lack of) prediction, I found very high consistency of results in the present data. Consider first the 12 criteria listed in Table 3. That table represents 60 independent factor-criterion evaluations (five factors \times 12 criteria), of which 18 were judged to have replicated across the four data sets with significant predictions (as shown in the 18 rows of the table).

Table 3
Replicated Partial Correlations (With Sex Removed) Between Criteria and Big Five Factor Scales

| Criterion | Factor scale | FF-NPQ (<i>n</i> = 178) | FF-NPQ (<i>n</i> = 142) | NEO-FFI (<i>n</i> = 273) | NEO-PI-R (<i>n</i> = 135) |
|------------------------------|-------------------|-----------------------------|-----------------------------|------------------------------|-------------------------------|
| Participant sex ^a | Neuroticism | .53*** | .54*** | .31*** | .20* |
| | Agreeableness | .33*** | .23** | .25*** | .23** |
| | Conscientiousness | .19** | .30*** | .06 | .22* |
| Intelligence | Neuroticism | -.23** | -.12 | -.30*** | -.33*** |
| | Openness | .13* | .24** | .25*** | .32*** |
| | Conscientiousness | .07 | .17* | .20*** | .36*** |
| Alcohol consumption | Extraversion | .30*** | .33*** | .14* | .24** |
| | Conscientiousness | -.21*** | -.15 | -.26*** | -.29*** |
| Honesty | Agreeableness | .23*** | .19* | .17** | .26** |
| | Conscientiousness | .25*** | .12 | .27*** | .27*** |
| Popularity | Extraversion | .39*** | .36*** | .39*** | .51*** |
| Parties attended | Extraversion | .31*** | .33*** | .29*** | .33*** |
| Grade point average | Conscientiousness | .23*** | .13 | .27*** | .19* |
| Plays musical instruments | Openness | .19** | .24** | .19** | .20* |
| Attractiveness | Extraversion | .25*** | .07 | .35*** | .18* |
| Tobacco consumption | Agreeableness | -.23*** | -.16 | -.26*** | -.18* |
| Dating variety | Extraversion | .17** | .24** | .07 | .28*** |
| Routinely exercises | Extraversion | .23*** | .18* | .29*** | .14 |

Note. FF-NPQ = Five-Factor Nonverbal Personality Questionnaire; NEO-FFI = NEO Five-Factor Inventory; NEO-PI-R = Revised NEO Personality Inventory.

^a Participant sex was not partialled.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Of the remaining 42 evaluations (60 – 18), 37 replicated across data sets with nonsignificant predictions (i.e., 88%). Thus, in only 5 cases (12%) were there some noteworthy disagreements among the four personality assessments in terms of whether a particular factor predicted a particular criterion. What of the predictions of the other 15 criteria not listed in Table 3? Their absence from that table means that none of those criteria met my standard of replicated significant predictions. For the standard of replicated null predictions, however, the large majority of the results for those 15 criteria showed consistency. Of the 75 evaluations of replicated null effects (five factors \times 15 criteria), 67 of them agreed in their results across the different Big Five assessments (89%).

Nonreplicated Big Five Predictors

The present evaluations of the predictive generalizability of Big Five factor measures across different inventories have shown substantially positive results overall, in terms of both replicated significant predictions and replicated null predictions. But there may be an additional informative value in studying the discrepant findings, in which some Big Five factor measures predicted a criterion but others did not. The specific discrepancies in factor-criterion predictions found in this study, based on correlations that failed the above-mentioned tests for replicated significant and replicated nonsignificant results, are listed in Table 4. The 13 behavior criteria listed in that table have been organized to highlight possible reasons for the discrepant findings in those particular cases.

Note the pattern of factor-criterion relations for the four personality assessments and the first seven criteria listed in Table 4. In each of those cases, the verbal NEO-FFI and NEO-PI-R

questionnaires resulted in the same factor predicting the same criterion, whereas the nonverbal FF-NPQ measures failed to replicate those results. This pattern of correlations suggests that there is possibly some difference in the content of the verbal and nonverbal scales that gives one an advantage over the other in the prediction of certain criterion variables. As an example, the NEO-PI-R Neuroticism scale contains statements intended to measure self-consciousness, referring to a tendency to suffer feelings of inferiority or discomfort when being scrutinized by others. Such behaviors are difficult to convey with nonverbal personality items of the type used in this study and, therefore, are not represented in the FF-NPQ Neuroticism measure. This might explain why the verbal Neuroticism scale but not the nonverbal scale predicted (low) self-perceived attractiveness.

The next four criteria listed in Table 4 (Rows 8 to 11) show systematic discrepancies in factor-criterion correlations across the two independent samples of participants. Specifically, Agreeableness predicted religiosity, and Openness to Experience predicted preference for liberal arts study in the smaller sample of participants ($n = 142$) but not in the larger sample ($n = 276$). The opposite sample effects were true for Agreeableness predicting dieting behavior and driving fast, both being significant (and negative) in the larger sample but not in the smaller sample. These results suggest that the participants in the two samples may have had slightly different personality compositions or different underlying behavior determinants.

The discrepant correlations for the last two factor-criterion combinations listed in Table 4 (Rows 12 and 13) are not interpretable in terms of differences in the verbal versus nonverbal personality assessments. Nor are they interpretable in terms of Sample 1 versus Sample 2 participant differences. It is likely that those

Table 4
Nonreplicated Partial Correlations (With Sex Removed) Between Criteria and Big Five Factor Scales

| Criterion | Factor scale | FF-NPQ (<i>n</i> = 178) | FF-NPQ (<i>n</i> = 142) | NEO-FFI (<i>n</i> = 273) | NEO-PI-R (<i>n</i> = 135) |
|------------------------------|-------------------|-----------------------------|-----------------------------|------------------------------|-------------------------------|
| Attractiveness | Neuroticism | .01 | -.08 | -.25*** | -.23** |
| Dating frequency | Conscientiousness | -.02 | .10 | .14* | .21* |
| Fraternity interest | Extraversion | .11 | .14 | .16* | .19* |
| Participation in sports | Extraversion | .10 | .12 | .13* | .26** |
| Routinely exercises | Neuroticism | -.13 | .03 | -.19** | -.18* |
| Traffic violations | Conscientiousness | -.03 | .02 | -.14* | -.20* |
| Participant sex ^a | Extraversion | -.08 | -.08 | .14* | .28** |
| Religiosity | Agreeableness | .08 | .25** | .00 | .19* |
| Liberal arts | Openness | .06 | .25** | .10 | .38*** |
| Dieting behavior | Agreeableness | -.15* | .05 | -.13* | .04 |
| Driving fast | Agreeableness | -.15* | -.17 | -.14* | -.17 |
| Tobacco consumption | Extraversion | .18** | .13 | -.09 | .22* |
| Dating variety | Agreeableness | -.19** | -.15 | -.10 | -.24** |

Note. FF-NPQ = Five-Factor Nonverbal Personality Questionnaire; NEO-FFI = NEO Five-Factor Inventory; NEO-PI-R = Revised NEO Personality Inventory.

^a Participant sex was not partialled.

* $p < .05$. ** $p < .01$. *** $p < .001$.

differences in prediction results are simply due to random measurement error.

Discussion

My purpose in this study was to evaluate the utility and generalizability of different assessments of the Big Five factors of personality as predictors of diverse behavior criteria. The criteria were argued to have wide-ranging social and cultural implications and included behaviors such as donating blood and driving automobiles at excessive speeds. These criterion variables are most likely complex variables in the sense that they are multifaceted and multidetermined—they do not simply represent alternative measures of the Big Five factors. Such alternative personality factor measures should, of course, correlate with the present Big Five predictors, but such evidence would not satisfy the purpose of this study regarding the evaluation of measurement utility and generalizability.¹

The results of this research generally support the interpretation that the present assessments of the Big Five factors of personality—using the NEO-FFI, NEO-PI-R, and a new five-factor measure called the FF-NPQ—are able, in many cases, to predict important behavior criteria. Furthermore, for the personality measures evaluated in this research, there was a remarkable consistency in their ability to predict those criteria. That consistency manifested itself in factor–criterion correlations that were well replicated, both in their significant effects and in their null effects.

Some further replication of the present personality–behavior correlations can be found in related data presented by Paunonen and Ashton (2001a). In that study, our participants ($N = 141$) also completed both the NEO-PI-R and the present version of the Behavior Report Form, among other measures. I have reviewed those data from the point of view of the 18 replicated partial correlations shown in Table 3. Of those 18 tests of predictor–criterion association, 13 were also significant ($p < .05$) in

Paunonen and Ashton's data (i.e., 72%), and all 18 were in the same direction as those shown in Table 3.

Validity of the Behavior Measures

The behavior criterion variables used in this study were derived from participants' self-ratings on individual items contained in the Behavior Report Form. One might legitimately question, therefore, the psychometric properties of those measures. The fact that they constitute single-item indicators of behavior suggests that they might be fraught with measurement error and, as a consequence, be of low reliability. Moreover, the fact that they constitute self-reports of behavior suggests that they might not agree with more objective behavior assessments and, as a result, might not be valid.

I have argued in the past that criterion ratings such as those used in this study represent mostly innocuous self-reports of behavior occurrence or frequency and, therefore, are probably relatively reliable and valid (e.g., Paunonen, 1998; Paunonen & Ashton, 2001a). For example, it seems reasonable to trust a person's self-report of whether he or she wears eyeglasses, is enrolled in a liberal arts program of study, or works at a part-time job. Some of the ratings, however, might be less trustworthy, because they involve a greater degree of subjective judgment and, thus, could be

¹ One might argue in this context that the criterion variables labeled as intelligence, femininity, religiosity, and honesty are merely narrow personality traits that each underlie some Big Five personality factor and, thus, ought to be predicted by those factors. However, although intelligence is expected to be related to the Openness to Experience factor, that individual-differences variable is not considered to be a facet of Openness or any other Big Five dimension (Allik & Realo, 1997; Costa & McCrae, 1992, p. 55). Furthermore, masculinity–femininity, religiosity, and honesty have been empirically determined to be three behavior domains, among several others, that are also questionable as facets of the Big Five factor space (see Ashton et al., 2000; Paunonen & Jackson, 2000).

prone to memory errors or desirability biases. In particular, the first six Behavior Report Form items might be problematic in this regard—the items that represent self-ratings of attractiveness, intelligence, popularity, masculinity–femininity, religiosity, and honesty.

Because of a well-founded concern about the psychometric properties of the present behavior criteria, empirical evidence of their validities is of some interest. One can estimate such validities by, for example, correlating the participants' self-reports on the behavior items with the reports of independent observers. I was, in fact, able to do this in the present study using data collected previously by Paunonen and Ashton (2001a). In that earlier study, we had participants living in a university residence ($N = 141$) complete the Behavior Report Form in its usual self-rating format. We then asked each participant's roommate to complete that questionnaire in a peer rating format, providing an independent report of each person's behavior.

The correlations between self-ratings and roommate ratings on each Behavior Report Form item, based on Paunonen and Ashton's (2001a) data, are listed in Table 5, ordered by magnitude. (Correlations could not be computed for three of the Behavior Report Form items—fraternity membership, use of diabetes medication, and use of depression medication—because of a lack of response variance.) The validity estimates for the Behavior Report Form items ranged in size from .12 to .92, with the mean of .52. For three of the four computed behavior criteria used in this study, the self–peer correlations were as follows: preference for contact lenses = .90, mean number of traffic violations = .49, and total medication usage = .45. (The computed obesity index could not be evaluated for validity because roommate judgments of weight and height were not collected.)

The sizes of the self–peer correlations for the present behavior ratings gain perspective if one compares them with typical results for personality measures. The mean correlation of .52 shown in Table 5 is as high as the .52 mean convergent validity reported by Paunonen and Jackson (1985, Table 2) between self-ratings and roommate ratings on the trait scales of the Personality Research Form (PRF; Jackson, 1984). The .52 mean value for the Behavior Report Form items is even higher than the .43 mean self–peer correlation reported for the domain scales of the NEO–PI–R (Costa & McCrae, 1992, Table 8). Unlike the present single-item measures, however, the PRF trait scales are 16-item measures, and the NEO–PI–R domain scales are 48-item measures. Validity estimates for those personality scales, therefore, benefit from the aggregation of items. What is the size of the correlation between self and peer for only a single item on the PRF or the NEO–PI–R? If one uses the Spearman–Brown formula (Lord & Novick, 1968, p. 114) and assumes that the self-rated and peer-rated personality scales have reliabilities of .70 for the PRF and .90 for the NEO–PI–R (close to those scales' values reported in their respective manuals), the expected correlation between self-ratings and peer ratings on any single personality item is only .10 for the PRF and .09 for the NEO–PI–R. As is evident in Table 5, the single items of the Behavior Report Form well exceeded this region of measurement validity (a notable exception being the .12 correlation between self-ratings and peer ratings of attractiveness).

Note in Table 5 that the relatively high correlations between self-reports and peer reports on some variables are not unexpected, such as whether one wears glasses ($r = .86$) and contact lenses

Table 5
Self–Peer Correlations for Behavior Report Form Items

| Item no. | Item label | r |
|----------|---------------------------|-----|
| 9a | Tobacco consumption | .92 |
| 9b | Number of cigarettes | .92 |
| 19c | Wears contacts | .90 |
| 19b | Wears glasses | .86 |
| 18a | Asthma medication | .84 |
| 20a | Part-time work | .73 |
| 10b | Number of drinks | .70 |
| 19a | Eyesight | .69 |
| 7 | Grade-point average | .66 |
| 21c | Routinely exercises | .65 |
| 10a | Alcohol consumption | .64 |
| 11 | Liberal arts | .60 |
| 15 | Buys lottery tickets | .60 |
| 13b | Speeding tickets | .58 |
| 21b | Participation in sports | .58 |
| 14 | High school grade | .57 |
| 20b | Part-time hours | .56 |
| 21a | Plays musical instruments | .56 |
| 5 | Religiosity | .55 |
| 13a | Driving experience | .52 |
| 18b | Allergy medication | .45 |
| 21d | Blood donations | .42 |
| 8b | Dating variety | .41 |
| 12b | Fraternity interest | .39 |
| 13d | Other traffic violations | .38 |
| 8a | Dating frequency | .37 |
| 2 | Intelligence | .35 |
| 13c | Parking tickets | .34 |
| 13e | Driving fast | .33 |
| 17 | Dieting behavior | .32 |
| 4 | Femininity | .30 |
| 18e | Other medication | .30 |
| 6 | Honesty | .23 |
| 3 | Popularity | .21 |
| 16 | Parties attended | .20 |
| 1 | Attractiveness | .12 |
| 12a | Fraternity membership | |
| 18c | Diabetes medication | |
| 18d | Depression medication | |

Note. Empty cells indicate that the correlation was not computed for the item because of lack of response variance.

($r = .90$), works at a part-time job ($r = .73$), takes asthma medication ($r = .84$), or participates in sports ($r = .58$). But there are other variables with high validities that are, perhaps, somewhat surprising. Those variables include ratings of alcohol and tobacco consumption ($r = .64$ and $.92$, respectively) and estimates of number of drinks imbibed per week ($r = .70$) and cigarettes smoked per day ($r = .92$). One might expect self-ratings of those less-than-desirable habits to show somewhat lower consensus with peer ratings, because of a penchant for some people to minimize the presence of negative attributes in rating themselves, especially in an experimental context, but not necessarily in rating others.

As surmised previously, self-ratings requiring more subjective judgments were not as highly correlated with the corresponding peer ratings. Specifically, ratings of attractiveness, popularity, honesty, femininity, and intelligence were near the bottom of the list in Table 5, with a mean estimated validity of only .24. The reason for their lower overall validity is probably that self-ratings on those variables are subject to self-enhancement biases that are

not part of peer ratings. This is why it is important to emphasize that it was the self-perception of intelligence, for example, that was the variable of interest in the results reported in this study, and not intelligence as inferred by others or as measured more objectively (which helps to explain the negative correlations between intelligence and Neuroticism observed in Table 3). The same interpretation applies to the present analyses of the attractiveness, popularity, honesty, and femininity criteria. It is interesting that the religiosity self-ratings seem to have had a less subjective self-enhancement component to them, showing a relatively high convergence of .55 with peer ratings.²

In general, the self-peer correlations for the present criteria indicated behaviors about which independent observers showed good consensus in their ratings of occurrence or frequency. And as long as one acknowledges some of the self-assessments for what they are—namely, self-perceptions, and not objective measures, of attractiveness, popularity, honesty, femininity, and intelligence—then it is maintained that the Behavior Report Form items are generally accurate indicators of behavior tendencies, having utility for behavior prediction research.

Validity of the FF-NPQ

I consider the consistency in behavior predictions found in this study to be remarkable mostly because of fundamental differences between the measures of the Big Five personality factors used. The FF-NPQ is substantially different in its items and format from the NEO-FFI and NEO-PI-R, despite the fact that the latter questionnaires were referred to in the construction of the former (see the FF-NPQ section in the introduction). The FF-NPQ is an experimental nonverbal personality questionnaire, whereas the NEO-FFI and the NEO-PI-R are standard verbal personality inventories. From this perspective, one would expect the two verbal forms to be much more similar to each other in their predictions than would either verbal form be to the nonverbal form. This is especially true of the present verbal assessments, because the items of the NEO-FFI are also the very same items found in the NEO-PI-R. However, the operations of all three questionnaires in predicting the criteria were quite similar. With the exception of a few cases described in Table 4, there was little evidence that the verbal forms were, in general, more similar to each other in predicting criterion variables than was either form to the nonverbal inventory.

The results of this study not only support the claim that there are significant relations between the Big Five personality factors and important behavior criteria but also support the construct validity of the new nonverbal measure of those factors, the FF-NPQ. (Simultaneously and equally, the results support the construct validity of the NEO-FFI and the NEO-PI-R.) Two sources of evidence in this study can be interpreted as substantiating the assertion that the measured FF-NPQ variables reside well within the five-factor model of personality. First, there were the relatively high correlations between the FF-NPQ scales and the corresponding verbal scales of the NEO-FFI and the NEO-PI-R. Those verbal inventories represent well-established, construct valid personality measures in their own right. The second source of evidence verifying the assertion that the FF-NPQ scales measure the Big Five personality factors exists in the congruence of the nonverbal and verbal prediction results vis-à-vis the present behavior criteria. If the nonverbal items were measuring substantially dif-

ferent factors than were the items of the two verbal forms, their empirical prediction results would be expected to be noticeably different, which they most certainly were not. Some corroborating evidence of convergent validity between FF-NPQ and NEO-FFI measurements and criterion predictions has been reported by Paunonen et al. (2001).³

As mentioned in the introduction, a nonverbal personality questionnaire has certain advantages over verbal measures of the same constructs. These advantages pertain primarily to the measurement of personality in different language groups. The fact that the questionnaire's items do not need to be translated into those languages means that cross-cultural research can be expedited. This is not to imply, however, that the FF-NPQ is a culture-free test. Its nonverbal items, in fact, are most relevant to educated, middle-class people in Western cultures. For detailed discussions of these and other issues related to using personality tests, both verbal and nonverbal, developed in one culture with people from another culture, see Paunonen (2000), Paunonen et al. (2001), and Paunonen and Ashton (1998, 2002).

Attenuated Levels of Prediction

One might be tempted to consider the levels of factor-criterion correlation found in this study to be unsubstantial, the significant ones being mostly in the range of .20 to .30. One should remember, however, the various causes that tend to lower correlations of the type reported in this article. First, despite the relatively high validities of the present criterion measures (Table 5), both those criteria and the personality predictors scales still contain some amount of measurement error. That measurement error serves to attenuate the observed factor-criterion interrelations. Attenuation-corrected predictive validity coefficients would necessarily be higher than the correlation coefficients reported in this article.

Another reason why the correlations in this study are not higher is because the criterion behaviors, as I have already stated, are

² We also had in our earlier study (Paunonen & Ashton, 2001a) university transcripts of the roommates' grades. Their self-ratings of grade point averages on the Behavior Report Form correlated well with their university records ($r = .47$) but not as highly as with their peers' rated grades ($r = .66$). The difference in those two correlations could be explained by a self-enhancement bias on the part of the students both in their reporting of grade point averages on the Behavior Report Form and in their recounting of those grades in conversations with their roommates. Those roommates would generally have to rely on their partners' verbal reports in making their peer ratings, having little access to more objective grade records.

³ Verbal mediation is a process that might be relevant in explaining the similarity between the present verbal and nonverbal personality assessments. That is, when completing the FF-NPQ, respondents could have attached verbal labels to the behavior acts represented in the nonverbal items. Their FF-NPQ item ratings could then have been based on those verbal cues, cues that might be similar to the verbal information used to arrive at their NEO-FFI and NEO-PI-R item responses. But even if verbal mediation did occur on the nonverbal questionnaire items, it is still remarkable that the verbal and nonverbal results were so similar. The extra encoding step involved in making a nonverbal item response would be expected to add to unreliability in the nonverbal scale scores, because not everyone would be using the same encoding process and consequent verbal labels in arriving at their item responses (see Paunonen & Jackson, 1979, for further discussion of verbal mediation in responding to the present type of nonverbal personality item).

probably multifaceted and multidetermined. This is to imply that Big Five personality factors are not the only variables involved in the determination of complex criteria such as grade point average, alcohol consumption, or tendency to engage in dieting behavior. Surely such behaviors have multiple causes, including genetic predispositions, hormonal influences, various environmental determinants, and their interactions. Personality may, in fact, have only a minor influence on the behaviors evaluated in this research. The empirical factor–criterion correlations would be expected to improve to the extent that other relevant predictors are included in the analyses.

A third reason why the correlations in this study are probably attenuated concerns the personality predictors themselves. Data have been presented elsewhere showing that, although the Big Five factors of personality have some utility in predicting criterion behaviors of the type studied here, the constituent facets of those factors can do better in that regard (e.g., Ashton, 1998; Ashton et al., 1995; Mershon & Gorsuch, 1988; Paunonen, 1993, 1998; Paunonen et al., 1999; Paunonen & Ashton, 2001a, 2001b; Paunonen & Nicol, 2001; Rothstein et al., 1994). Consider the example of low Conscientiousness predicting alcohol consumption shown in Table 3. This (replicated) empirical finding might not make theoretical sense at first glance. However, Conscientiousness is a factor of personality that represents several distinct but correlated personality traits or facets. These facets purportedly include orderliness, achievement orientation, and self-discipline, among others. The aforementioned negative correlation could thus be entirely due to an inverse association between self-discipline and alcohol consumption. It might have absolutely nothing to do with orderliness, achievement orientation, or any other facet of that factor. Consequently, combining several relatively narrow Conscientiousness facets together into a single indicator of that broad personality factor could actually lead to a lower coefficient of prediction than would be obtained if the individual facets were treated as separate predictors.

There is yet another reason for the seemingly low coefficients of correlation reported between the present personality variables and behavior criteria. There may be other variables, personality variables, that might be implicated in the prediction of the behaviors but that were not included in this study. I refer not to the facets of the Big Five personality factors mentioned in the preceding paragraph but to other lower level personality traits that do not fit well into the five-factor model of personality structure and, hence, are not incorporated as components of typical Big Five factor measures. I have studied such variables (Paunonen & Jackson, 2000) and concluded that there might be several well-defined personality traits that exist beyond the Big Five factors (cf. Saucier & Goldberg, 1998). Honesty is one possible example of such a dimension (see also Ashton, Lee, & Son, 2000). Including a measure of honesty in the present study might have increased the prediction of some of the behavior criteria, such as number of traffic violations, for instance. Other non-Big Five variables might have served as useful predictors of other criteria.

There are many reasons why the present levels of factor–criterion correlation might be as high as can be expected under the circumstances. Changing those circumstances could readily lead to nontrivial improvements in the accuracies of behavior predictions. Adding different personality variables into the predictor set would be an easy first step. Less easy would be the identification and

assessment of relevant nonpersonality determinants of behavior, such as genetic factors, environmental influences, and their interplay.

Conclusions

The present results are interpreted as supporting the following conclusions. The Big Five factor inventories used in this study show a strong convergence in their measurements and predictions. That consistency of results supports the construct validity of those inventories. Furthermore, the replicated behavior predictions found in this research strongly implicate specific personality constructs as underlying determinants of specific complex behaviors. That conclusion notwithstanding, it is argued that even better empirical predictions can be achieved if other variables are included in the battery of behavior predictors, leading to an even better understanding of the nomological network of behavior determinants. Those other variables might include (a) the narrow personality facets that constitute the broad Big Five factors, (b) personality traits that are largely independent of the Big Five, and (c) other determinants of behavior related to genetics and the environment.

References

- Allik, J., & Realo, A. (1997). Intelligence, academic abilities, and personality. *Personality and Individual Differences, 23*, 809–814.
- Ashton, M. C. (1998). Personality and job performance: The importance of narrow traits. *Journal of Organizational Behavior, 19*, 289–303.
- Ashton, M. C., Jackson, D. N., Paunonen, S. V., Helmes, E., & Rothstein, M. G. (1995). The criterion validity of broad factor scales versus specific trait scales. *Journal of Research in Personality, 29*, 432–442.
- Ashton, M. C., Lee, K., & Son, C. (2000). Honesty as the sixth factor of personality: Correlations with Machiavellianism, primary psychopathy, and social adroitness. *European Journal of Personality, 14*, 359–368.
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1–26.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.
- De Raad, B., & Perugini, M. (Eds.). (2002). *Big Five assessment*. Göttingen, Germany: Hogrefe & Huber.
- Digman, J. M. (1990). Personality structure: Emergence of the Five-Factor Model. *Annual Review of Psychology, 41*, 417–440.
- Funder, D. C. (2001). Personality. *Annual Review of Psychology, 52*, 197–221.
- Goff, M., & Ackerman, P. L. (1992). Personality–intelligence relations: Assessment of typical intellectual engagement. *Journal of Educational Psychology, 84*, 537–552.
- Hough, L. M. (1992). The “Big Five” personality variables–construct confusion: Description versus prediction. *Human Performance, 5*, 139–155.
- Jackson, D. N. (1984). *Personality Research Form manual*. Port Huron, MI: Research Psychologists Press.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Menlo Park, CA: Addison-Wesley.
- McCrae, R. R., & John, O. P. (1992). An introduction to the Five-Factor Model and its applications. *Journal of Personality, 60*, 175–215.

- Mershon, B., & Gorsuch, R. L. (1988). Number of factors in the personality sphere: Does increase in factors increase predictability of real-life criteria? *Journal of Personality and Social Psychology*, *55*, 675–680.
- Paunonen, S. V. (1993, August). *Sense, nonsense, and the Big Five factors of personality*. Paper presented at the 101st Annual Convention of the American Psychological Association, Toronto, Ontario, Canada.
- Paunonen, S. V. (1998). Hierarchical organization of personality and prediction of behavior. *Journal of Personality and Social Psychology*, *74*, 538–556.
- Paunonen, S. V. (2000). Construct validity and the search for cross-situational consistencies in personality. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 123–140). Norwell, MA: Kluwer.
- Paunonen, S. V., & Ashton, M. C. (1998). The structured assessment of personality across cultures. *Journal of Cross-Cultural Psychology*, *29*, 150–170.
- Paunonen, S. V., & Ashton, M. C. (2001a). Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology*, *81*, 524–539.
- Paunonen, S. V., & Ashton, M. C. (2001b). Big Five predictors of academic achievement. *Journal of Research in Personality*, *35*, 78–90.
- Paunonen, S. V., & Ashton, M. C. (2002). The nonverbal assessment of personality: The NPQ and the FF–NPQ. In B. De Raad & M. Perugini (Eds.), *Big Five assessment* (pp. 171–194). Göttingen, Germany: Hogrefe & Huber.
- Paunonen, S. V., Ashton, M. C., & Jackson, D. N. (2001). Nonverbal assessment of the Big Five personality factors. *European Journal of Personality*, *15*, 3–18.
- Paunonen, S. V., & Jackson, D. N. (1979). Nonverbal trait inference. *Journal of Personality and Social Psychology*, *37*, 1645–1659.
- Paunonen, S. V., & Jackson, D. N. (1985). Idiographic measurement strategies for personality and prediction: Some unredeemed promissory notes. *Psychological Review*, *92*, 486–511.
- Paunonen, S. V., & Jackson, D. N. (2000). What is beyond the Big Five? Plenty! *Journal of Personality*, *68*, 821–835.
- Paunonen, S. V., Jackson, D. N., & Keinonen, M. (1990). The structured nonverbal assessment of personality. *Journal of Personality*, *58*, 481–502.
- Paunonen, S. V., Jackson, D. N., Trzebinski, J., & Forsterling, F. (1992). Personality structure across cultures: A multimethod evaluation. *Journal of Personality and Social Psychology*, *62*, 447–456.
- Paunonen, S. V., Keinonen, M., Trzebinski, J., Forsterling, F., Grishenko-Roze, N., Kouznetsova, L., & Chan, D. W. (1996). The structure of personality in six cultures. *Journal of Cross-Cultural Psychology*, *27*, 339–353.
- Paunonen, S. V., & Nicol, A. A. M. (2001). The personality hierarchy and the prediction of work behaviors. In R. Hogan & B. W. Roberts (Eds.), *Personality psychology in the workplace* (pp. 161–191). Washington, DC: American Psychological Association.
- Paunonen, S. V., Rothstein, M. G., & Jackson, D. N. (1999). Narrow reasoning about the use of broad personality measures in personnel selection. *Journal of Organizational Behavior*, *20*, 389–405.
- Paunonen, S. V., Zeidner, M., Engvik, H., Oosterveld, P., & Maliphant, R. (2000). The nonverbal assessment of personality in five cultures. *Journal of Cross-Cultural Psychology*, *31*, 220–239.
- Rothstein, M. G., Paunonen, S. V., Rush, J. C., & King, G. A. (1994). Personality and cognitive ability predictors of performance in graduate business school. *Journal of Educational Psychology*, *86*, 516–530.
- Saucier, G., & Goldberg, L. R. (1998). What is beyond the Big Five? *Journal of Personality*, *66*, 495–524.
- Tett, R. P., Jackson, D. N., & Rothstein, M. G. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, *44*, 703–742.

Appendix

Behavior Report Form

INSTRUCTIONS

Please **describe yourself** by answering all of the questions in this booklet. Be as accurate as possible in describing you, and try not to omit any item.

First, please indicate the following about you.

Your Name: _____

Your Student No.: _____

Your Sex: (1) ___ M or (2) ___ F

Your Age: ___ (years)

Your Height: ___ (in.) or ___ (cm.)

Your Weight: ___ (lbs.) or ___ (kg.)

Your Hand Preference: (1) ___ Right or (2) ___ Left

Your Year of Study: _____

Please answer the following six questions by **circling a number** on each of the corresponding 9-point rating scales.

1. How would you rate your level of *physical attractiveness* compared to the average student?

Circle a number on the following scale:

| | | | | | | | |
|--------------|------|------------|----------|------------|------------|------|-----------|
| UNATTRACTIVE | | | AVERAGE | ATTRACTIVE | | | |
| 1 | 2 | 3 | 5 | 6 | 7 | 8 | 9 |
| Extremely | Very | Moderately | Somewhat | Somewhat | Moderately | Very | Extremely |

2. How would you rate your level of *general intelligence* compared to the average student?

Circle a number on the following scale:

| | | |
|---|----------------|-------------------------------|
| UNINTELLIGENT | AVERAGE | INTELLIGENT |
| 1 2 3 4 5 6 7 8 9 | | |
| Extremely Very Moderately Somewhat | Somewhat | Moderately Very Extremely |

3. In relation to people you know, how would you describe your *popularity* among your peers?

| | | |
|---|----------------|-------------------------------|
| UNPOPULAR | AVERAGE | POPULAR |
| 1 2 3 4 5 6 7 8 9 | | |
| Extremely Very Moderately Somewhat | Somewhat | Moderately Very Extremely |

4. In relation to other people of your sex, how *masculine/feminine* are you?

| | | |
|---|-------------------------------|-------------------------------|
| MASCULINE | AVERAGE FOR MY SEX | FEMININE |
| 1 2 3 4 5 6 7 8 9 | | |
| Extremely Very Moderately Somewhat | Somewhat | Moderately Very Extremely |

5. Would you consider yourself a *religious* person who, for example, might be interested in attending (or already does attend) formal religious services?

| | | |
|---|----------------|-------------------------------|
| NONRELIGIOUS | AVERAGE | RELIGIOUS |
| 1 2 3 4 5 6 7 8 9 | | |
| Extremely Very Moderately Somewhat | Somewhat | Moderately Very Extremely |

6. How would you describe your *honesty* compared to that of your friends?

| | | |
|---|----------------|-------------------------------|
| DISHONEST | AVERAGE | HONEST |
| 1 2 3 4 5 6 7 8 9 | | |
| Extremely Very Moderately Somewhat | Somewhat | Moderately Very Extremely |

7. Indicate your *overall average grade* for last year. ____ % (percent)

8. Estimate the average number of *dates per month* that you have had over the past year with someone of the opposite sex? ____ (per month)

How many *different people* did you date over the course of the past year? ____

9. Do you consider yourself a *nonsmoker, infrequent smoker, light smoker, moderate smoker, or heavy smoker*? Circle a number below.

| | | | | |
|----------------|----------------------|-----------------|--------------------|-----------------|
| 1 | 2 | 3 | 4 | 5 |
| Non- Smoker | Infrequent Smoker | Light Smoker | Moderate Smoker | Heavy Smoker |

If you smoke, *how many cigarettes* do you smoke a day on average? ____ (per day)

10. Do you consider yourself a *nondrinker, infrequent drinker, light drinker, moderate drinker, or heavy drinker* of alcohol? Circle a number below.

| | | | | |
|-----------------|-----------------------|------------------|---------------------|------------------|
| 1 | 2 | 3 | 4 | 5 |
| Non- Drinker | Infrequent Drinker | Light Drinker | Moderate Drinker | Heavy Drinker |

If you drink, *how many drinks* do you consume a week on average?
(one drink = 1 ounce spirits = 1 glass wine = 1 bottle beer) ____ (drinks per wk.)

11. If you had to choose one program of study (or, if you have already chosen one), which of the two below would you (did you) choose? Check one.

- (1) ____ *Liberal Arts* (Arts, Fine Arts, Humanities, etc.)
(2) ____ *Non-Liberal Arts* (Natural Sciences, Social Sciences, Professional Schools, etc.)

(Appendix continues)

12. Are you currently a *fraternity/sorority member* or pledge? Check one.

(1) ___ no (2) ___ yes

If you are not currently a fraternity/sorority member or pledge, *how much interest* would you have in joining a fraternity/sorority?

| | | |
|---------------------|----------------|-------------------|
| UNINTERESTED | NEUTRAL | INTERESTED |
| 1 | 5 | 9 |
| 2 | | 8 |
| 3 | | 7 |
| 4 | | 6 |
| Extremely | Somewhat | Moderately |
| Very | Somewhat | Very |
| Moderately | | Extremely |

13. *How long* have you had a driver's license? Put 0 if never. ___ (years)

If you have a driver's license:

How many *speeding tickets* have you had in the past? ___

How many *parking tickets*? ___

How many *other* traffic violations? ___

What is the *fastest* you have driven? ___ (m.p.h.) or ___ (k.p.h.)

14. Indicate your *graduating average grade* from high school. ___ % (percent)

15. On average, *how much money* do you spend monthly buying lottery tickets?
___ (\$ per month)

16. Estimate the average number of *parties per month* that you attend. ___ (per month)

17. Have you ever been on a *long-term diet* (i.e., one month or more)? Check one.

(1) ___ no (2) ___ yes

18. Do you regularly take any *prescription medications* for the following ailments? Check each item.

asthma: (1) ___ no (2) ___ yes

allergies: (1) ___ no (2) ___ yes

diabetes: (1) ___ no (2) ___ yes

depression: (1) ___ no (2) ___ yes

other: (1) ___ no (2) ___ yes

19. How would you describe your *uncorrected eyesight*? Check one.

(1) ___ normal vision

(2) ___ nearsighted

(3) ___ farsighted

(4) ___ nearsighted in one eye, farsighted in the other

Do you wear vision-correcting *prescription lenses*? Check both items.

glasses: (1) ___ no (2) ___ yes

contacts: (1) ___ no (2) ___ yes

20. Do you currently hold a *part-time job*? Check one.

(1) ___ no (2) ___ yes

If yes, *how many hours* do you work in the average week? ___ (hrs. per week)

21. Do you: play any *musical instrument(s)*? (1) ___ no (2) ___ yes
 play any *organized sport(s)*? (1) ___ no (2) ___ yes
exercise regularly (aerobics, weights, running, etc.)? (1) ___ no (2) ___ yes
donate blood occasionally? (1) ___ no (2) ___ yes

Note. From "Hierarchical organization of personality and prediction of behavior," by S. V. Paunonen, 1998, *Journal of Personality and Social Psychology*, 74, 538–556. Copyright 1998 by the American Psychological Association.

Received June 19, 2001
 Revision received June 21, 2002
 Accepted July 2, 2002 ■