NEO Five-Factor Inventory Scores: Psychometric Properties in a Community Sample

Greg Murray, David Rawlings, Nicholas B. Allen, and John Trinder

Psychometric analysis of scores from the NEO Five-Factor Inventory (NEO-FFI) confirmed their internal (N = 527) and test–retest reliability (n = 335) in an Australian adult sample. The data also provided support for the internal and temporal reliability of scores derived from 13 lower order item clusters identified in the NEO-FFI by G. Saucier (1998).

Over the last decade, a consensus has emerged that five dimensions might be sufficient both to cover personality space and to represent the deductions from major theoretical approaches (Goldberg, 1993). This view has come to be known as the five-factor model (FFM) of personality (for a review, see O'Connor, 2002). The vast research output associated with Costa and McCrae's (1992) NEO Personality Inventory-Revised (NEO-PI-R) has led some researchers to identify this particular formulation with the FFM (Digman, 1990). The five domains of personality as described by Costa and McCrae are labelled Neuroticism (N), Extraversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C). The NEO formulation has proven useful in clinical research (e.g., Groth-Marnat & Jeffs, 2002; Rector, Hood, Richter, & Bagby, 2002) and has been explored as a tool in clinical practice (e.g., McCrae & Costa, 1991; Miller, 1991).

The NEO Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1992) is an abbreviated 60-item version of the 240-item NEO-PI-R. As described by Costa and McCrae, the NEO-FFI was constructed by first selecting the 12 items with the largest structure coefficients for each of the five factors. Subsequently, 10 substitutions were made to permit reverse keying of some items in each scale, diversify item content, and eliminate items with joint coefficients. In a sample of 1,539 employees of a large national corporation, Costa and McCrae found coefficients of internal consistency ranging from .68 (A) to .89 (N).

Because of its brevity, comprehensiveness, and ease of administration, the NEO-FFI has great promise as both a clinical and a research tool (Tokar, Fischer, Snell, & Harik-Williams, 1999). However, beyond the original samples investigated by Costa and McCrae (1992), the psychometric properties of scores from the NEO-FFI have not been widely tested. It has been argued that researchers and others are overly optimistic about generalizing psychometric findings from parent instruments to their associated short forms (Smith, McCarthy, & Anderson, 2000), and the NEO-FFI has yet to receive the psychometric attention that it warrants.

Exploratory factor analyses have been the focus of three studies. Using an availability sample of 243 female college students, Holden and Fekken (1994) concluded that the latent structure of the instrument was consistent with its scoring key and that the internal consistency of the five scales was adequate. Exploratory factor analyses identified only five items that were misplaced (i.e., primary pattern coefficients were not on the scale to which they belonged). It is worth noting that the study of Holden and Fekken did not meet sample size criteria for factor analysis ($N > 300$; Tabachnick & Fidell, 1996).

Greg Murray, School of Social and Behavioural Science, Swinburne University of Technology, Hawthorn, Victoria, Australia; David Rawlings, Nicholas B. Allen, and John Trinder, Department of Psychology, University of Melbourne, Victoria, Australia. *The data presented here are drawn from a longitudinal study undertaken as part of Greg Murray's doctoral research. The information and analyses in no way duplicate other publications from the study. The authors would like to thank Annie Fiume, Dominic Murray-Fiume, and Angelica Murray-Fiume for their assistance with the research.* *Correspondence concerning this article should be addressed to Greg Murray, School of Social and Behavioural Science, Swinburne University of Technology, Hawthorn, Victoria, Australia (e-mail: gwm@swin.edu.au).*

In a sample of 485 nonstudent adults, Tokar and colleagues (Tokar et al., 1999) interpreted three, four-, five-, and six-factor solutions. Both the five- and six-factor solutions represented the five personality factors well. The six-factor solution explained slightly more variance (33% vs. 31%), identified the FFM more clearly, and also identified a sixth interpretable factor (measuring "likeability" or "consideration"). Consistent with Holden and Fekken's (1994) analyses, the authors concluded that their data supported the conventional scoring of the NEO-FFI and demonstrated its practical utility as a FFM indicator.

Egan and colleagues (Egan, Deary, & Austin, 2000) analyzed the psychometrics of the inventory among 1,025 adult participants pooled across three divergent British samples. In contrast to the two aforementioned reports, nine items were found to be misplaced. Using the stricter criterion of relatively unique items relating to particular traits, Egan et al. identified 14 problematic items and concluded that the NEO-FFI requires modification (Egan et al., 2000, p. 907).

Data on the test–retest reliability of scores from the NEO-FFI have been reported in only one sample, as described by Costa and McCrae (1992). A subset ($n = 208$) of a college sample that completed the NEO-PI-R for norming purposes had also completed the NEO-FFI 3 months previously. By selecting appropriate items from the NEO-PI-R data, temporal reliability of the five NEO-FFI scale scores could be estimated. Derived coefficients were .79, .79, .80, .75, and .83 for N, E, O, A, and C, respectively ($p < .001$ in each case). There are two limitations to these estimates. First, time periods of less than 1 year can be considered short term for assessing stability in personality reports (Schuerger, Zarrella, & Hotz, 1989). For example, stability in life events over this period may artificially inflate test–retest correlations. Second, the true stability of personality may be lower among college students than among older adults (Roberts, Caspi, & Moffitt, 2001).

The NEO-FFI provides only five broad domain scores and not the lower level facet scores that are measured on the NEO-PI-R. However, using a split-sample replication strategy, Saucier (1998) has identified internally reliable lower order item clusters from the NEO-FFI. Separate factor analyses of the 12 items of each of the five scales identified 13 clusters that fit within the FFM's broad-bandwidth factors (N: Negative Affect, Self-Reproach; E: Positive Affect, Sociability, Activity; O: Aesthetic Interests, Intellectual Interests, Unconventionality; A: Nonantagonistic Orientation, Prosocial Orientation; C: Orderliness, Goal-Striving, Dependability). As argued by Saucier, the use of the NEO-FFI would be significantly expanded if these lower level scale scores proved reliable in replication samples.

THE PRESENT STUDY

Much remains unknown about the psychometric properties of the NEO-FFI. In the present research, we aimed to advance understanding in three ways. First, previous research provides ambiguous support for the factor structure and internal reliability of scores from the instrument. We addressed this issue in the present study in a randomly drawn sample of the adult population. Second, we replicated the suggestion that 13 lower order facets may also be reliably measured on the instrument. Third, test–retest reliability data has only been reported on a small college sample over a short time period, and the Saucier (1998) scale scores have not yet been subjected to temporal reliability analysis. In the present research, we took repeated measurements of NEO-FFI scores across 30 months.

METHOD

Design

Personality data were collected as part of a six-wave panel study using a mail-out survey methodology. As described elsewhere (Murray, Allen, Rawlings, & Trinder, 2002; Murray, Allen, & Trinder, 2001), the study was designed to investigate personality correlates of seasonal variation in mood. The present study focuses on the psychometrics of the personality measure.

Sampling and Participants

Potential participants were sampled randomly from the electoral roll in a suburban district of Melbourne, Australia. The locale was selected for its high rates of home ownership and English as a first language and is consequently a homogeneously white-collar district

(Victorian Electoral Commission, 1999). One thousand and eighty names were assigned to two batches. As a control for possible sequence effects, Batch 1 only completed a preliminary questionnaire (T_{prelim}), the data from which was not used in the present analysis. Both batches then completed questionnaires in August and February across 3 years (Time 1 [T1]–Time 6 [T6]). Analyses were based on T1–T6 data collapsed across both batches. Participants were removed from the mailing list if they failed to respond at any wave or if they moved to reside outside the locale of the study.

As shown in Table 1, 530 individuals responded at T1, of whom 63.2% ($n = 335$) completed all subsequent waves of the study. At T1, the mean age was 46.9 years ($SD = 17.0$), and 293 members (55.3%) of the sample were female. The majority of participants (83.5%) were born in Australia. A substantial minority (39.3%) reported having dependent children. A comparison with census data for the region found that individuals around the age of 50 were overrepresented in the present sample, while individuals in their mid-20s were relatively underrepresented. The distribution of both gender and country of birth were comparable with census data for the district. Completion of the full six waves was associated with a number of features at T1, namely, higher C, $t(343.5) = -3.93, p < .01$, $d = .42$; higher A, $t(339.5) = -2.85, p < .001$, $d = .31$; and older age, $t(365.6) = -4.12, p < .001$, $d = .43$. Among the 335 respondents at T6 (54.9% female), the mean age was 49.4 ($SD = 16.0$).

Measures

The NEO-FFI was completed at each wave, along with a number of mood scales that are not relevant to the present investigation. Age, gender, and country of birth were recorded to permit comparison of the derived sample with census data for the region. The NEO-FFI was scored using the standard system. As recommended in the manual (Costa & McCrae, 1992), missing data were replaced with the neutral response, and questionnaires with 10 or more missing data points were considered invalid.

Procedure

A number of methods were used to maximize response rate (see Dillman, 1978). These included mailing up to three personally addressed letters at each wave, a monetary incentive included with the questionnaires, and a between-wave courtesy letter.

Analyses

T1 provided the largest sample on which to conduct item-level analyses. Exploratory factor analyses were conducted on T1 responses to investigate the latent structure of the in-

TABLE 1

Number of Responses and Response Rate by Wave and Number of Questionnaires Sent and Returned at Each Wave by Batch

Batch	T_{prelim}	T1	T2	T3	T4	T5	T6
1							
Sent	500	269	232	218	199	181	169
Returned	270	238	218	200	181	169	161
2							
Sent	—	580	286	249	218	198	181
Returned	—	292	249	220	199	182	174
Total response per wave	270	530	467	420	380	351	335
Response rate percentage	56.0	62.4	90.2	89.9	91.1	92.6	95.7

Note. T_{prelim} = Time zero questionnaire completed by Batch 1 only to control for possible sequence effects; T1 = Time 1; T2 = Time 2; T3 = Time 3; T4 = Time 4; T5 = Time 5; T6 = Time 6.

ventory (both orthogonal and oblique rotations were investigated). Extraction of factors was based on principal components analysis of correlation matrices. Inspection of the scree plot of eigenvalues was used to determine the number of factors. This was supported by the parallel analysis method of Horn (1965), the most consistently accurate of all the methods for determining the number of factors (Zwick & Velicer, 1986). Delta was set at 0 for the oblique rotation solution. In line with previous research, interpretation of pattern coefficients was based on two criteria: the factor on which the item had its largest coefficient and the factor(s) on which coefficients were significant (> |.30|). Internal reliability of scores from the five scales was investigated using coefficient alpha, with values greater than .7 being considered adequate (Nunnally, 1978).

The lower level latent structure of the NEO-FFI was explored by investigating the 13 subscales proposed by Saucier (1998). The sample size was not sufficient to provide a meaningful replication of Saucier's original findings concerning the within-scale factor structure of the instrument. We therefore adopted the strategy of assuming Saucier's factor solution and investigating the internal reliability of the 13 subscale scores in the present data set.

RESULTS

Distribution of FFM Scores and Comparison With Norms

Of 530 respondents at T1, 3 (0.6%) were excluded from analyses because they had missing data on 10 or more items of the NEO-FFI. Item-level analyses were therefore performed on a sample of 527 participants. As shown in Table 2, the distributions of NEO factor scores in the present sample were comparable with published U.S. norms (Costa & McCrae, 1992). Being female was associated with higher scores on N and A, a trend that can also be seen in the U.S. data.

Factor Structure

Orthogonal rotation. An exploratory principal components analysis using orthogonal rotation was conducted on the 60 items of the NEO-FFI. Initial analysis of the correlation matrix identi-

TABLE 2

Means and Standard Deviations of the Five NEO Five-Factor Inventory (NEO-FFI) Scales at T1 (N = 527) and Normative Data (Norms) in a U. S. Sample (N = 1,539; Costa & McCrae, 1992)

	All Participants		Men		Women	
Sample	*M*	*SD*	*M*	*SD*	*M*	*SD*
N						
T1	17.8	8.2	16.6[a]	7.9	18.8[a]	8.3
Norms	19.1	7.7	17.6	7.5	20.5	7.6
E						
T1	28.3	6.6	28.2	6.7	28.4	6.5
Norms	27.7	5.9	27.2	5.9	28.2	5.8
O						
T1	28.7	6.6	28.1	6.7	29.2	6.5
Norms	27.0	5.8	27.1	5.8	27.0	5.9
A						
T1	32.6	5.6	31.3[b]	5.5	33.7[b]	5.5
Norms	32.8	5.0	31.9	5.0	33.8	4.7
C						
T1	34.1	6.5	33.7	6.6	34.4	6.5
Norms	34.6	5.9	34.1	6.0	35.0	5.8

Note. N = Neuroticism; E = Extraversion; O = Openness to Experience; A = Agreeableness; C = Conscientiousness; T1 = Time 1.

[a]Male-female differences $p < .01$, $d = .27$. [b]Male-female differences $p < .001$, $d = .44$.

fied 13 factors with eigenvalues greater than unity. The first six of these appeared before an elbow in the scree plot, and parallel analysis confirmed that six factors should be retained. The six factors explained 41.7% of the total variance (14.5%, 7.8%, 6.5%, 5.5%, 4.4%, and 3.1%, respectively).

An orthogonal rotation was first carried out on these six factors. Although this is not reported in detail, we note that the first five factors of this rotation closely paralleled the five factors of the five-factor rotation (noted later in this discussion). A sixth factor was produced whose largest pattern coefficients were from Openness to Experience items associated with liberal values (Item 38, Item 18) and the desire for novel experience (Item 28, Item 8). In addition, the factor included a number of Extraversion items scored in the direction of introversion and pointing to the desire for independence and autonomy (Item 57, Item 27).

Because the aim of this study was to explore the NEO-FFI in terms of the FFM, the first five factors were then retained for rotation. These five explained 38.7% of the total variance. Rotation of the five factors to orthogonal simple structure using Varimax produced a set of pattern coefficients that delineated the FFM structure. As shown in Table 3, 57 of the 60 items had primary and significant coefficients on the expected factor. The 12 N scale items had primary coefficients on Factor 1, the 12 E scale items had primary coefficients on Factor 3, and the 12 C scale items had primary coefficients on Factor 2.

Of the 12 O scale items, 11 had primary pattern coefficients on Factor 4, but Item 8 had its primary coefficient on C and only a secondary coefficient on O. Ten of the 12 A scale items had their primary coefficient on Factor 5. Item 29 had its primary coefficient on N and a secondary one on A. Item 34 had its highest coefficient on E and a secondary one on A. The three misplaced items, therefore, all had secondary coefficients on the correct scale, but the magnitude of the secondary coefficient varied from adequate to weak (.38, .25, and .19 for Items 29, 34, and 8, respectively).

The internal reliability alphas for scores from the a priori N, E, O, A, and C scales were .87, .80, .77, .75, and .85, respectively. Reliability analysis suggested that the five scales were not compromised by the three items with aberrant primary coefficients. Removal of Item 8 from the O scale did not increase alpha substantially (.768 to .774). Removal of either of the aberrant items from the A scale marginally decreased alpha for the scale (from .748 to .739 [Item 29 removed] and .744 [Item 34 removed]).

Oblique rotation. An oblique rotation (oblimin) was then applied to the first five factors to see if this better represented the latent structure of the NEO-FFI. Examination of the resultant pattern matrix again offered strong support for the instrument's scoring key. Fifty-seven of the 60 items showed primary and substantial pattern coefficients on their corresponding components. The three misplaced items found using orthogonal rotation (discussed previously) were also misplaced in the same way using oblique rotation, namely, Item 8 had its primary coefficient on C, Item 29 had its primary coefficient on N, and Item 34 had its primary coefficient on E. The similarity between the oblique and orthogonal rotations was not surprising in light of the low correlations between factors in the oblique analysis (–.03 to .18).

Internal reliability of Saucier's (1998) lower level scale scores. The internal reliability of scores derived from the 13 lower level item clusters was adequate. Internal reliability alpha was substantial for both of the N facets (.73, .83), the three E facets (.71, .60, .68), the three O facets (.74, .74, .44), the three C facets (.75, .69, .68), and somewhat lower for the two A facets (.68, .56). These reliabilities are comparable with those reported by Saucier (1998) for a replication sample ($N = 253$). In both the present sample and Saucier's replication group, the N facet Self-Reproach (7 items) recorded the highest alpha (.83 in both studies), and the O facet Unconventionality (4 items) recorded the lowest alpha (.44 and .40, respectively). The mean of the 13 alphas in the present sample was .68 and in Saucier's replication sample was .66.

Test–Retest Reliability

As summarized in Table 4, the test–retest reliabilities for all five scale scores tended to decrease as the retest interval lengthened but were nonetheless robust across 30 months. Six-month reliabilities ranged from .80 (A) to .87 (O), with a mean correlation across scale scores of .83 ($SD = .03$). Medium-term (30-month) reliabilities ranged from .73 (A) to .86 (O; $M = .79$, $SD = .05$).

TABLE 3

Rotated Component Matrix for Five Retained Factors After Principal Components Analysis of the 60 NEO Five-Factor Inventory (NEO-FFI) Scale Items

NEO-FFI Item Number	Factor 1 (N)	Factor 3 (E)	Factor 4 (O)	Factor 5 (A)	Factor 2 (C)
N Scale					
1	**.45**	−.10	.12	.04	.14
6	**.61**	−.15	−.09	.05	−.16
11	**.63**	−.08	.09	−.04	−.14
16	**.58**	−.04	.14	−.02	−.07
21	**.69**	−.11	.00	−.13	−.07
26	**.68**	−.11	−.01	−.06	−.19
31	**.64**	−.10	.07	.02	−.02
36	**.60**	−.02	−.04	−.39	−.07
41	**.64**	−.09	−.07	−.09	−.27
46	**.60**	−.16	.09	.01	−.03
51	**.64**	−.12	−.15	.06	−.29
56	**.55**	−.07	.07	−.12	−.14
E Scale					
2	−.01	**.62**	−.10	.10	−.14
7	−.14	**.58**	.13	.04	−.01
12	−.19	**.47**	.07	.04	−.15
17	−.05	**.55**	.22	.24	.01
22	.01	**.60**	−.05	−.14	.16
27	−.18	**.37**	.09	.13	−.07
32	−.09	**.49**	.21	−.11	.25
37	−.12	**.71**	.02	.05	.16
42	−.38	**.42**	.17	.12	.12
47	.00	**.53**	.11	−.18	.21
52	−.01	**.47**	.09	.01	.44
57	−.20	**.43**	.02	−.02	.14
O Scale					
3	.17	−.15	**.36**	.07	−.32
8	−.12	−.08	.19	.04	**−.47**
13	.09	.04	**.63**	.07	−.02
18	−.18	.08	**.48**	−.06	.03
23	.07	.03	**.63**	.04	−.06
28	−.06	.30	**.33**	−.08	−.08
33	.13	.05	**.49**	.09	.09
38	−.07	.01	**.32**	−.01	−.08
43	.10	.03	**.70**	.05	−.01
48	.06	.05	**.72**	.09	.04
53	.01	.27	**.65**	−.11	.18
58	.02	.10	**.62**	−.20	.02
A Scale					
4	.10	.23	.03	**.48**	.14
9	−.25	−.05	−.02	**.43**	.23
14	−.12	.07	−.15	**.61**	.13
19	.10	.01	.06	**.41**	−.02
24	−.28	.13	.08	**.49**	.06
29	**−.43**	.04	.23	.38	−.12
34	−.17	**.44**	−.03	.25	.08
39	−.04	.21	−.05	**.66**	−.03
44	−.05	−.17	.06	**.56**	−.23
49	.16	.26	.17	**.46**	.21
54	−.12	−.05	.00	**.58**	−.04
59	−.06	−.19	−.09	**.52**	.17
C Scale					
5	−.05	−.01	−.15	.09	**.54**
10	−.16	−.05	−.07	.03	**.67**

(continued on next page)

TABLE 3 (Continued)

Rotated Component Matrix for Five Retained Factors After Principal Components Analysis of the 60 NEO Five-Factor Inventory (NEO-FFI) Scale Items

NEO-FFI Item Number	Factor 1 (N)	Factor 3 (E)	Factor 4 (O)	Factor 5 (A)	Factor 2 (C)
C Scale *(continued)*					
15	−.21	−.16	.06	−.04	**.55**
20	.06	.01	.06	.29	**.59**
25	−.19	.13	.00	−.12	**.64**
30	−.28	−.04	−.17	−.03	**.52**
35	−.07	.26	.13	.02	**.68**
40	−.04	.03	.01	.32	**.53**
45	−.21	−.03	−.07	.24	**.44**
50	−.09	.22	.05	.07	**.68**
55	−.32	.02	−.07	−.04	**.62**
20	.03	.16	.14	−.05	**.57**
Eigenvalue	5.72	5.48	4.32	3.99	3.69
Percent variance	9.53	9.14	7.20	6.65	6.15

Note. N = Neuroticism; E = Extraversion; O = Openness to Experience; A = Agreeableness; C = Conscientiousness. The largest pattern coefficient is shown in bold.

Results contained in Table 4 show that the test–retest reliabilities of scores from the proposed 13 lower order scales were also adequate. Six-month reliabilities ranged from .68 (Prosocial Orientation) to .81 (Aesthetic Interests), with a mean correlation across scales of .75 ($SD = .04$). As with the temporal reliability of factor scores, correlations decreased steadily over longer time intervals but nevertheless at 30 months ranged from .62 (Prosocial Orientation) to .79 (Positive Affect; $M = .72$, $SD = .05$).

DISCUSSION

A substantial sample of randomly identified adults in Melbourne, Australia, completed the NEO-FFI on six occasions. At T1, the data from 527 of 530 respondents was available for analysis. Compared with census data for the region, the sample was somewhat biased toward older age groups but was an adequate representation of the region in terms of gender and country of birth. At T6, 335 participants remained in the sample. Not surprisingly, completing all six waves of the study was positively associated with T1 levels of C, A, and age. The distributions of the five NEO dimensions in the sample at T1 were generally comparable with published norms.

When constrained to a five-factor orthogonal solution, item-level analyses of T1 data correctly located all but three of the 60 items using orthogonal rotation. One of the misplaced items (Item 8: "When I find the right way to do something, I stick to it" [reverse-scored]) from the Action facet of Openness appeared to operate primarily as a measure of C. Two items from the A scale were misplaced. Item 29 ("I believe that most people will take advantage of you if you let them" [reverse-scored]) from the Altruism facet of A, appeared to be operating primarily as a measure of N. Item 34 ("Most people I know like me") is from the Trust facet of A but seems to be operating primarily as a measure of E in this data set.

Each of these three items has been identified as problematic in previous research. Consistent with the current array of pattern coefficients, Egan et al. (2000) found that Item 8 is primarily a measure of low C. In the research of both Egan et al. and Holden and Fekken (1994), Item 29 had its primary coefficient on N, and Item 34 had its primary coefficient on E. The discrepant coefficients appear to be intelligible in light of the FFM. Conscientiousness was conceived by Costa and McCrae (1992) as either an inhibitory dimension

TABLE 4

Test–Retest Reliabilities (From 6–30 Months) for Scores From the Five NEO Five-Factor Inventory (NEO-FFI) Scales and Saucier's (1998) 13 Lower Order Scales

Scale	T1–T2 (6 Months, *n* = 462)	T1–T3 (12 Months, *n* = 419)	T1–T4 (18 Months, *n* = 377)	T1–T5 (24 Months, *n* = 350)	T1–T6 (30 Months, *n* = 334)
NEO-FFI scale scores					
N	.80	.78	.79	.77	.75
E	.86	.84	.80	.82	.83
O	.87	.87	.86	.85	.86
A	.80	.81	.80	.74	.73
C	.85	.83	.80	.79	.80
Saucier's lower order scale scores					
Negative Affect (N)	.71	.69	.70	.67	.71
Self-Reproach (N)	.77	.76	.76	.73	.73
Positive Affect (E)	.77	.76	.74	.73	.79
Sociability (E)	.78	.79	.76	.75	.74
Activity (E)	.79	.76	.73	.74	.74
Aesthetic Interests (O)	.81	.81	.81	.78	.78
Intellectual Interests (O)	.77	.79	.78	.78	.75
Unconventionality (O)	.72	.71	.70	.69	.71
Nonantagonistic Orientation (A)	.76	.79	.78	.72	.69
Prosocial Orientation (A)	.68	.68	.70	.64	.62
Orderliness (C)	.81	.78	.77	.74	.74
Goal-Striving (C)	.75	.73	.71	.68	.69
Dependability (C)	.70	.69	.69	.66	.67

Note. N = Neuroticism; E = Extraversion; O = Openness to Experience; A = Agreeableness; C = Conscientiousness. T1 = Time 1; T2 = Time 2; T3 = Time 3; T4 = Time 4; T5 = Time 5; T6 = Time 6.

("prudence" or "constraint") or a proactive dimension ("will to achieve"). It is not surprising, then, that "When I find the right way to do something, I stick to it" has a prominent coefficient on C. Similarly, given that Hostility and Self-Consciousness are facets of N on the NEO, it is reasonable that the item "I believe that most people will take advantage of you if you let them" has a substantial pattern coefficient on N. Likewise, "Most people I know like me" might be expected to generate a substantial pattern coefficient on E, which is defined by facets such as Gregariousness and Warmth.

These three misplacements, although apparently robust across samples and having intelligible associations with other Big Five traits, may not severely compromise the validity of the test scores. In the present sample, each of the three misplaced items exhibited secondary pattern coefficients on the expected factor, and in no case did internal reliability analyses suggest that the items' deviation from full-scale scores warranted their exclusion from the scale. Consistent with previous studies, scores from all five scales demonstrated internal reliability against the criterion of $\alpha > 0.70$. Indeed, as noted by an anonymous reviewer, the relationships observed between NEO-FFI items in this contemporary Australian sample are remarkably similar to those found in the culturally distinct sample (Baltimore circa 1986; see Costa & McCrae, 1992) used to develop the instrument.

Investigation of the proposed lower order scales of the NEO-FFI (Saucier, 1998) found that scores from the 13 item clusters had adequate internal reliability. Furthermore, the pattern of internal reliabilities across the scale scores was very similar to that observed by Saucier.

Previous studies (Egan et al., 2000; Holden & Fekken, 1994; Tokar et al., 1999) have drawn dissimilar conclusions about the internal reliability of factor scores from the NEO-FFI. Our findings from the present study are the most positive to date, with only three items found to be misplaced in both orthogonal and oblique solutions. The most defensible conclusion at this stage is that, on balance, the factor structure of the NEO-FFI appears adequate.

Test–retest reliabilities for each of the five NEO scale scores were significant and large effects across administrations up to 30 months. The derived 6-month reliability estimates were of similar magnitude to previously published short-term reliability figures (Costa & McCrae, 1992). Reliability decreased slightly over time, but the medium-term (30-month) reliabilities remained substantial. This finding is in accord with existing data on personality stability (McCrae, 1994). Test–retest reliabilities for scores from the 13 lower order scales were of a similar order of magnitude. Adding to Saucier's (1998) original work, these data therefore provide further evidence that the NEO-FFI can be reliably used to measure the FFM at a more specific level than the five broad domains. Further research into the reliability and validity of these novel facets appears warranted.

It has been argued that researchers and others are overly optimistic about generalizing psychometric findings from parent instruments to their associated short forms (Smith et al., 2000). The NEO-FFI is an important short form, the use of which is supported by the present psychometric analysis. As an efficient source of reliable personality scores, important theoretical deductions from the FFM can usefully be tested by counselors (e.g., Sanderson & Clarkin, 1994) and novel clinical hypotheses explored. Saucier's (1998) 13 lower order scales seem to offer particular promise to the clinician. The ability to distinguish between, for example, Goal-Striving and Orderliness as elements of client Conscientiousness may constitute a substantial advance in formulation and intervention planning.

CONCLUSION

The present findings add weight to the claim that the NEO-FFI is an effective brief measure of the Big Five model of personality and shows promise as a measure of lower order facets. There is a broader debate in personality research about the validity of the FFM, and alignments in this debate may have partly influenced the standards that previous studies have set for the NEO-FFI. The present research had a tight focus on the psychometrics of the inventory's scores, and the data collected here suggest that they are internally and temporally reliable.

REFERENCES

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory and NEO Five-Factor Inventory professional manual*. Odessa, FL: Psychological Assessment Resources.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41,* 417–440.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method*. New York: Wiley.

Egan, V., Deary, I., & Austin, E. (2000). The NEO-FFI: Emerging British norms and an item-level analysis suggest N, A, and C are more reliable than O and E. *Personality and Individual Differences, 29,* 907–920.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48,* 26–34.

Groth-Marnat, G., & Jeffs, M. (2002). Personality factors from the five-factor model of personality that predict dissociative tendencies in a clinical population. *Personality & Individual Differences, 32,* 969–976.

Holden, R. R., & Fekken, G. C. (1994). The NEO Five-Factor Inventory in a Canadian context: Psychometric properties for a sample of university women. *Personality and Individual Differences, 17,* 441–444.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179–185.

McCrae, R. R. (1994). Moderated analyses of longitudinal personality stability. *Journal of Personality and Social Psychology, 65,* 577–585.

McCrae, R. R., & Costa, P. T., Jr. (1991). The NEO personality inventory: Using the five-factor model in counseling. *Journal of Counseling & Development, 69,* 362–372.

Miller, T. R. (1991). The psychotherapeutic utility of the five-factor model of personality: A clinician's experience. *Journal of Personality Assessment, 57,* 415–433.

Murray, G., Allen, N. B., Rawlings, D., & Trinder, J. (2002). Seasonality and personality: A prospective investigation of Five Factor Model correlates of mood seasonality. *European Journal of Personality, 16,* 457–468.

Murray, G., Allen, N. B., & Trinder, J. (2001). A longitudinal investigation of seasonal variation in mood. *Chronobiology International, 18,* 875–891.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
O'Connor, B. P. (2002). A quantitative review of the comprehensiveness of the five-factor model in relation to popular personality inventories. *Assessment, 9,* 188–203.
Rector, N., Hood, K., Richter, M., & Bagby, R. (2002). Obsessive-compulsive disorder and the five-factor model of personality: Distinction and overlap with major depressive disorder. *Behaviour Research & Therapy, 40,* 1205–1219.
Roberts, B. W., Caspi, A., & Moffitt, T. E. (2001). The kids are alright: Growth and stability in personality development from adolescence to adulthood. *Journal of Personality and Social Psychology, 81,* 670–683.
Sanderson, C., & Clarkin, J. F. (1994). Use of the NEO-PI personality dimensions in differential treatment planning. In P. T. Costa & T. A. Widiger (Eds.), *Personality disorders and the five-factor model of personality* (pp. 219–235). Washington, DC: American Psychological Association.
Saucier, G. (1998). Replicable item-cluster subcomponents in the NEO Five-Factor Inventory. *Journal of Personality Assessment, 70,* 263–276.
Schuerger, J. M., Zarrella, K. L., & Hotz, A. S. (1989). Factors that influence the temporal stability of personality by questionnaire. *Journal of Personality and Social Psychology, 56,* 777–783.
Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment, 12,* 102–111.
Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: HarperCollins.
Tokar, D. M., Fischer, A. R., Snell, A. F., & Harik-Williams, N. (1999). Efficient assessment of the Five-Factor Model of personality: Structural validity analyses of the NEO Five-Factor Inventory (Form S). *Measurement and Evaluation in Counseling and Development, 32,* 14–30.
Victorian Electoral Commission. (1999). *Information kit for 1999–2000 Victorian state election.* Melbourne, Australia: Author.
Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99,* 432–442.