JOURNAL OF PERSONALITY ASSESSMENT, *81*(3), 272–281

Exploring Issues of Personality Measurement and Structure Through the Development of a Short Form of the Eysenck Personality Profiler

K. V. Petrides
School of Psychology and Human Development
Institute of Education
University of London

Chris J. Jackson
School of Psychology
University of Queensland

Adrian Furnham
Department of Psychology
University College London

Stephen Z. Levine
Department of Psychology
University of Surrey

In this article, we develop a revised short form of the original Eysenck Personality Profiler (EPP; H. J. Eysenck & Wilson, 1991). In addition, we address topics of broad theoretical importance such as the recurrent empirical finding of correlations between conceptually orthogonal personality dimensions and the possibility that gender differences in these dimensions are partly spurious. In Study 1 (N = 227), we demonstrate that the existing short form of the EPP (EPP–SF; H. J. Eysenck, Wilson, & Jackson, 1996) provides a poor fit to the data and we develop a revised well-fitting version. In Study 2, we retest this version on an independent new sample (N = 3,374) where it is again found to fit the data well. We show that most of the structural and measurement parameters of the revised EPP–SF are invariant across genders. Structured means analysis indicated a significant gender difference in Psychoticism, with men scoring higher than women, but no differences in Extraversion or Neuroticism. Our discussion focuses on issues concerning personality measurement and structure, including an examination of the role of confirmatory factor analysis in personality research.

Most salient personality theories maintain that traits are hierarchically organized, with a few broad factors at the apex of the structure and a number of lower order factors (*primaries*) below (H. J. Eysenck & Eysenck, 1985; Goldberg, 1993; McCrae & Costa, 1999). Many such structures have been proposed (Cattell, 1973; Costa & McCrae, 1992a; H. J. Eysenck, 1992), differing both in the number of broad factors they postulate as well as in the substantive theories that underpin them.

Hans Eysenck was one of the major proponents of the hierarchical view of traits and the developer of one of the most influential theories of personality (H. J. Eysenck, 1947, 1990, 1997). The primary measurement instrument of Eysenckian personality is the Eysenck Personality Questionnaire–Revised (EPQ–R; S. B. G. Eysenck, Eysenck, & Barrett, 1985). The EPQ–R measures the three Eysenckian superfactors—namely, Psychoticism, Extraversion and Neuroticism—which have been found to be replicable across many different cultures and countries (Barrett, Petrides, Eysenck, & Eysenck, 1998).

In contrast to questionnaires that aim to assess primary factors (e.g., Cattell, Eber, & Tatsuoka, 1970), the EPQ–R is geared toward direct measurement of the superfactors. How-

ever, a combination of lower and higher level measurement as, for example, in the Revised NEO Personality Inventory (Costa & McCrae, 1992b), increases the breadth and detail of measurement and provides marker variables for the complex lower levels of trait structures.

In recognition of the advantages of combined measurement, H. J. Eysenck and colleagues (H. J. Eysenck, Barrett, Wilson, & Jackson, 1992; H. J. Eysenck & Wilson, 1991) developed a 21-scale inventory to measure finer aspects of the Eysenckian superfactors. The full version of the Eysenck Personality Profiler (EPP) comprises 420 items, that is, 20 for each of the 21 scales of the inventory.

The factor structure of the EPP has been the object of several psychometric investigations, most of which involved various forms of exploratory factor analysis (EFA; Costa & McCrae, 1995; H. J. Eysenck et al., 1992; Jackson, Furnham, Forde, & Cotter, 2000). These studies have suggested that some EPP scales either measure more than one superfactor (i.e., they are "factorially complex") or do not fit well into the Eysenckian personality hierarchy (but may fit into other hierarchies such as the Big Five; see Costa & McCrae, 1995).

The EPP was constructed for use in work-related settings and most of its applications focus on organizational psychology issues (e.g., job satisfaction; Furnham, Petrides, Jackson, & Cotter, 2002) or are based on employee samples (e.g., Jackson et al., 2000). However, the inventory can also be used in nonoccupational settings (e.g., Francis, Jones, Jackson, & Robbins, 2001; Wilson & Jackson, 1994). The EPP is a measure of normal adult personality, but, like other similar inventories, it may prove useful in the investigation of psychological disorders (e.g., Bienvenu et al., 2001). Nevertheless, no such applications have been carried out to date, despite the psychopathological origins of the Eysenckian dimensions of Neuroticism and Psychoticism.

STUDY 1

Because the full version of the EPP proved too lengthy for certain applications, H. J. Eysenck, Wilson, and Jackson (1996) developed a shorter version (EPP–SF) consisting of nine scales. Jackson et al. (2000) briefly examined the short form of the EPP through confirmatory factor analysis (CFA) and found that it did not provide an adequate fit to the data. Our main aim in the first study was to scrutinize the factor structure of the existing EPP–SF via CFA and to develop an improved new version.

Brief Conceptual Introduction to CFA

The main aim of CFA is to quantify, through a number of different indexes, the extent to which a data set is consistent with a factor structure that has been hypothesized a priori. Researchers do not explore the structure of their data after they have collected them, as in traditional EFA, but rather, they test the degree to which the data conform to an a priori structure. The degree of consistency between the sample data and the hypothesized structure can be quantified through a series of indexes such as the chi-square (χ^2), the comparative fit index (CFI), the standardized root mean square residual (SRMR), and the root mean square error of approximation (RMSEA).

Sometimes, it is erroneously assumed that CFA may be used if and only if researchers have a single, or a few, fixed a priori models that they wish to test on a particular sample. This strict interpretation of the potential uses of CFA is mistaken because for a long time it has been common to use the technique in a quasi-confirmatory manner, involving post hoc, data-driven modifications of the original a priori model (Jöreskog & Sörbom, 1993). Such modifications should be substantively interpretable and ideally cross-validated on independent samples (MacCallum, Roznowski, & Necowitz, 1992). A thorough and accessible discussion of applications of structural equation modeling, including CFA, can be found in MacCallum and Austin (2000). In this study, we used CFA in a quasi-confirmatory manner to derive a statistically optimal model for the short version of the EPP, which we cross-validated in Study 2.

Method

Participants. A total of 227 individuals participated in the study, of whom 125 were men, with a mean age of 37.9 years (SD = 6.3). All participants were employees in a transport company.

Measures and procedure. Participants completed the long form of the EPP from which the short form of the inventory can be derived. The existing EPP–SF comprises 9 of the 21 scales of the full version, 3 for each of the Eysenckian superfactors. Risk-Taking, Impulsivity, and Irresponsibility are marker scales for Psychoticism; Activity, Sociability, and Assertiveness are marker scales for Extraversion; and Inferiority, Unhappiness, and Anxiety are marker scales for Neuroticism. Each of these scales consists of 20 items responded to on a trichotomous scale (*yes*/*can't decide*/*no*).

The data were collected in the context of an assessment and development center. Item-level data were not available, and therefore it was not possible to estimate the internal consistencies of the scales on this sample. Although internal consistency estimates are sample specific (Thompson, 1994), the alphas for the EPP scales have consistently been shown to be adequate (e.g., Costa & McCrae, 1995; H. J. Eysenck et al., 1992; Jackson et al., 2000; see Table 4).

Results and Discussion

A maximum likelihood CFA was performed on the covariance matrix of the nine EPP–SF scales. The results indicated that this model did not fit the data well: $\chi^2(24, N =$

227) = 136.53; CFI = .80; SRMR = .12; and RMSEA = .13. Inspection of the modification indexes revealed the presence of many cross-loadings and error covariances between the nine scales. In view of the apparent lack of fit, we investigated an alternative set of EPP scales providing purer measurement of the Eysenckian dimensions. Extant factor pattern matrices from several studies (Costa & McCrae, 1995; H. J. Eysenck et al., 1992; Jackson et al., 2000) were inspected and a few sequentially modified models were empirically tested to arrive at a well-fitting solution.

The foregoing process culminated in a model with three indicators per superfactor and a cross-loading from a Psychoticism scale (Impulsivity) on the Neuroticism factor. The revised model provided a reasonable approximation to the sample data: $\chi^2(23, N = 227) = 64.64$; CFI = .92; SRMR = .06; and RMSEA = .09. The final scales included in the model, along with the standardized parameter estimates, are presented in Table 1. The revised model retained six of the existing EPP–SF scales (two per superfactor), but replaced Irresponsibility with Sensation-Seeking (Psychoticism [P]), Assertiveness with Ambitiousness (Extraversion [E]), and Inferiority with Dependence (Neuroticism [N]). The correlations between the three factors were $r_{PE} = .40$, $r_{NE} = -.38$, and $r_{PN} = .09$. Gender-specific statistics for theses scales are presented in Table 2.

To determine the relationship between the short and the long versions of the EPP, Levy's (1967) correction, which partials out the error variance that is common between the scores of the two forms, was applied.[1] The "true score" correlations, an index of strength of association between the factor scores derived from the long and short versions of the EPP, are presented separately for men and women in Table 3. Note that the true score correlations will always be lower than the corresponding zero-order correlations that include common error variability arising from item overlap.

A common feature shared by virtually all the scales that were dropped from the long form of the EPP is that they have been repeatedly found to load either on all three superfactors indiscriminately (e.g., Obsessive, Irresponsible, Aggressive, and Expressive) or on none at all (e.g., Practical; Costa & McCrae, 1995; H. J. Eysenck et al., 1992; Jackson et al., 2000; Moosbrugger & Fischbach, 2002). Note that some of the retained scales also seem to be complex factorially, both in this sample as well as in those of the aforementioned studies. For example, Impulsivity often tends to cross-load on Neuroticism, and even though the magnitude of this loading is usually small, the corresponding parameter was freed in this case because it was theoretically meaningful to do so, as we discuss later. Other possible cross-loadings, however, were not modeled because it was not as straightforward to justify them theoretically. For example, Sociability was not allowed to cross-load, despite the fact that it has consistently shown secondary and even tertiary loadings in other studies.

[1] A computer program for applying the Levy (1967) correction can be downloaded from Paul Barrett's homepage at http://www.liv.ac.uk/~pbarrett/paulhome.htm

TABLE 1
Completely Standardized Parameter Estimates and Factor Correlations for the Revised EPP–SF (Study 1)

	Factors			
Scales	*PSY*	*EXT*	*NEU*	*Uniqueness*
RIS	.69	.00	.00	.52
IMP	.51	.00	.29	.63
SEN	.73	.00	.00	.46
ACT	.00	.66	.00	.56
SOC	.00	.56	.00	.69
AMB	.00	.67	.00	.55
UNH	.00	.00	.81	.35
ANX	.00	.00	.78	.39
DEP	.00	.00	.74	.45

Note. $N = 227$. Factor correlations were $r_{PE} = .40$; $r_{NE} = -.38$; $r_{PN} = .09$. All parameter estimates were significant beyond the .05 level. EPP–SF = Short form of the Eysenck Personality Profiler; PSY = Psychoticism (P); EXT = Extraversion (E); NEU = Neuroticism (N); RIS = Risk-Taking; IMP = Impulsivity; SEN = Sensation-Seeking; ACT = Activity; SOC = Sociability; AMB = Ambitiousness; UNH = Unhappiness; ANX = Anxiety; DEP = Dependence.

In line with Jackson et al. (2000), the results of Study 1 showed that the original version of the EPP–SF (H. J. Eysenck et al., 1996) did not fit the data well. In contrast, the revised model provided a reasonable fit. Before further discussing these findings, it is necessary to establish, in light of the post hoc procedures used in this study, whether the proposed new model can provide an adequate approximation to data from an independent sample.

STUDY 2

The purpose of Study 2 was fourfold. First, it was to corroborate the model that was empirically derived in Study 1 (see Table 1) on an independent new sample. Second, we aimed to investigate the nature of EPP Impulsivity with reference to Dickman's (1990) distinction between two types of impulsivity, namely, functional and dysfunctional. The loadings of EPP Impulsivity on both Neuroticism and Psychoticism suggest that this scale taps pathological aspects of personality, and consequently, it should be expected to correlate more strongly with dysfunctional than with functional impulsivity.

The third objective was to examine whether the factor structure and factor loadings of the EPP–SF scales are invariant across the male and female samples. This is important because gender differences in the ways in which items are interpreted may in turn be manifested in differential factor loadings and/or factor variances and covariances. Finally, we aimed to investigate gender differences in the latent means of the three personality dimensions.

TABLE 2
Gender-Specific Descriptives and *t* Tests for Studies 1[a] and 2[b]

		Study 1					Study 2				
		Men[c]			Women[d]		Men[e]			Women[f]	
Factor and Scale	Descriptive Adjectives	M	Effect Sizes	SD	M	SD	M	Effect Size	SD	M	SD
Psychoticism											
RIS	Adventurous, daring, hesitant (R)	15.80	–.09	5.68	15.29	5.28	21.23**	.37	6.80	18.68	6.75
IMP	Hurried, impetuous, careful (R)	12.65	.03	6.33	12.45	6.13	17.10*	–.07	7.69	17.64	7.75
SEN	Excitement-seeking, disinhibited, subdued (R)	19.66*	.47	6.11	16.70	6.49	23.45**	.50	6.93	19.83	7.55
Extraversion											
ACT	Energetic, fact-paced, sluggish (R)	23.20	–.20	6.51	24.68	7.94	26.87*	.09	7.14	26.17	7.27
SOC	Talkative, outgoing, reserved (R)	25.84	–.21	6.80	27.25	6.27	26.70	.04	8.36	26.33	7.78
AMB	Competitive, hard-working, apathetic (R)	21.43	–.07	7.32	21.94	7.00	23.84*	.18	7.58	22.45	7.42
Neuroticism											
UNH	Gloomy, miserable, cheerful (R)	4.82	.13	5.27	4.19	4.17	6.67**	–.14	7.28	7.72	7.19
ANX	Nervous, worried, relaxed (R)	8.20	–.16	6.96	9.28	6.19	9.96**	–.34	7.49	12.61	7.69
DEP	Helpless, vulnerable, autonomous (R)	6.70	–.03	5.45	6.87	4.74	6.94**	–.19	5.97	8.12	5.81

Note. Effect sizes are based on Cohen's *d*. RIS = Risk-Taking; R = reverse scored; IMP = Impulsivity; SEN = Sensation-Seeking; ACT = Activity; SOC = Sociability; AMB = Ambitiousness; UNH = Unhappiness; ANX = Anxiety; DEP = Dependence.
[a]$N = 227$. [b]$N = 3{,}373$. [c]$n = 125$. [d]$n = 102$. [e]$n = 1{,}957$. [f]$n = 1{,}417$.
*Gender difference at $p < .05$. **Gender difference at $p < .01$.

TABLE 3
Gender-Specific and Total Sample "True Score" Correlations Between the Factors in the Long and Short Forms of the EPP for Study 1[a] and Study 2[b]

	Study 1			Study 2		
Factor	Men[c]	Women[d]	Total	Men[e]	Women[f]	Total
PSY	.65	.53	.60	.69	.70	.69
EXT	.63	.62	.62	.64	.64	.64
NEU	.87	.76	.83	.88	.86	.87

Note. RIS = Risk-Taking; IMP = Impulsivity; SEN = Sensation-Seeking; ACT = Activity; SOC = Sociability; AMB = Ambitiousness; UNH = Unhappiness; ANX = Anxiety; DEP = Dependence.
[a]$N = 227$. [b]$N = 3{,}373$. [c]$n = 125$. [d]$n = 102$. [e]$n = 1{,}957$. [f]$n = 1{,}417$.

Further Uses of CFA

In addition to conventional applications of CFA, which, as noted in Study 1, are aimed at testing a priori factor structures, it is also possible to use the technique to test for factor and latent mean structure invariance. Investigations of latent mean structure invariance aim to compare the factor means (i.e., average scores on the latent factors) of different groups or of a single group tested on different occasions.

Method

Participants. Data collected from various administrations of the long form of the EPP were collated for this study. The resultant sample comprised 3,374 individuals of whom 1,957 were men. The mean age for the sample was 32.74 years ($SD = 8.36$). The sample consisted mainly of adult employees from many different organizations and occupations, such as sales and recruitment consultants, information technology workers, and solicitors, but also included students and individuals from various social groups.

In addition, a supplementary sample comprising 44 undergraduate psychology students (38 women and 6 men) was employed to test the specific hypothesis about the nature of the Impulsivity scale in the EPP; 42 participants were between 18 and 25 years old and 2 were between 26 and 35.

Measures and procedure. The study was based on the revised version of the EPP–SF, which was developed in Study 1 (see Table 1). Item-level data were again unavailable, and therefore the internal consistencies of the scales could not be estimated. Table 4 shows the internal consistencies of the nine scales comprising the revised EPP–SF as reported in four previous studies employing the long form of the EPP.

Participants in the supplementary sample completed the long form of the EPP and Dickman's (1990) impulsivity scale, which comprises 23 items (11 measuring functional and 12 measuring dysfunctional impulsivity). Neither the data from Study 1 nor those from the supplementary sample in Study 2 were incorporated into the larger EPP data set ($N = 3{,}374$).

Results

Table 2 presents gender-specific descriptive statistics and *t* tests for the larger EPP data set. The revised EPP–SF model was tested through a multigroup (men and women) maximum likelihood CFA for cross-validation purposes as well as

TABLE 4
Internal Consistencies for the Revised EPP–SF Scales Based on Four Previous Studies

Factor and Scale	Costa & McCrae (1995)[b]	Eysenck et al. (1992)[a]		Jackson et al. (2000)[c]	Muris et al. (2000)[d]
		M	*F*		
Psychoticism					
RIS	.71	.69	.68	.66	.69
IMP	.79	.75	.75	.75	.76
SEN	.81	.75	.76	.74	.74
Extraversion					
ACT	.83	.75	.77	.71	.74
SOC	.84	.82	.81	.78	.75
AMB	.82	.80	.80	.77	.72
Neuroticism					
UNH	.87	.85	.89	.83	.88
ANX	.83	.83	.85	.80	.80
DEP	.63	.75	.77	.73	.71

Note. These alphas are based on studies that administered the long form of the Eysenck Personality Profiler (EPP). EPP–SF = short form of the EPP; Eysenck et al. = H. J. Eysenck, Barrett, Wilson, and Jackson; Jackson et al. = Jackson, Furnham, Forde, and Cotter; Muris et al. = Muris, Schmidt, Merckelbach, and Rassin; RIS = Risk-Taking; IMP = Impulsivity; SEN = Sensation-Seeking; ACT = Activity; SOC = Sociability; AMB = Ambitiousness; UNH = Unhappiness; ANX = Anxiety; DEP = Dependence. [a]*N* = 1,599; gender-specific data (M = males; F = females). [b]*N* = 229. [c]*N* = 655. [d]*N* = 215; Dutch data.

TABLE 5
Common Metric Completely Standardized Parameter Estimates and Factor Correlations for the EPP–SF (Study 2)

	Men[a]				*Women*[b]			
Scales	*PSY*	*EXT*	*NEU*	*Uniqueness*	*PSY*	*EXT*	*NEU*	*Uniqueness*
RIS	.78	.00	.00	.40	.78	.00	.00	.37
IMP	.60	.00	.38	.51	.62	.00	.23	.56
SEN	.69	.00	.00	.49	.76	.00	.00	.47
ACT	.00	.78	.00	.40	.00	.78	.00	.39
SOC	.00	.65	.00	.60	.00	.50	.00	.72
AMB	.00	.57	.00	.68	.00	.61	.00	.61
UNH	.00	.00	.86	.26	.00	.00	.86	.26
ANX	.00	.00	.77	.40	.00	.00	.72	.50
DEP	.00	.00	.85	.29	.00	.00	.81	.33

Note. *N* = 3,374. Factor correlations for men were $r_{PE} = .47$; $r_{NE} = -.46$; $r_{PN} = .02$; factor correlations for women were $r_{PE} = .46$; $r_{NE} = -.41$; $r_{PN} = -.03$. All parameter estimates were significant beyond the .05 level. EPP–SF = Short form of the Eysenck Personality Profiler; PSY = Psychoticism (P); EXT = Extraversion (E); NEU = Neuroticism (N); RIS = Risk-Taking; IMP = Impulsivity; SEN = Sensation-Seeking; ACT = Activity; SOC = Sociability; AMB = Ambitiousness; UNH = Unhappiness; ANX = Anxiety; DEP = Dependence.
[a]*N* = 1,957. [b]*N* = 1,417.

to establish a baseline model for the invariance and structured means analyses to follow. The model provided a reasonable approximation to the data—$\chi^2(46, N = 3,374) = 916.66$; CFI = .92; SRMR = .06; and RMSEA = .10—thus indicating that the post hoc procedures employed in Study 1 did not result in unacceptable levels of capitalization on chance. In Table 5, it can be seen that the common metric completely standardized parameter estimates for men and women were similar to those reported in Study 1. This was also the case for the gender-specific factor correlations.

The cross-loading of Impulsivity on Neuroticism was replicated on this sample, which reflects the pathological nature of EPP Impulsivity. Further evidence was provided by the fact that EPP Impulsivity was significantly correlated with Dickman's (1990) dysfunctional impulsivity, $r(44) = .56$, $p < .01$, but not with functional impulsivity, $r(44) = .10$, $p = ns$. The difference between these two correlations was statistically significant, $t(41) = 2.80$, $p < .01$.

The fit of the tested model indicated a common factor structure for men and women, but it did not speak to the issue of factor loadings invariance. Although Table 5 shows that the parameter estimates were similar across genders, one of the objectives of this study was statistically to test their equivalence. Accordingly, the entire factor pattern matrix was held invariant across genders, thereby creating a nested model to be tested through the $\Delta\chi^2$ test. The results indicated a significant fit attenuation, $\Delta\chi^2(7, N = 3,374) = 36.45$, $p < .01$.

In light of the significant overall chi-square change test, each element in the factor pattern matrix was tested individually through a sequence of cumulatively nested models. A summary of the results from this analysis is presented in Table 6. The loadings of Sociability on Extraversion and of Impulsivity on Neuroticism were significantly different in the male and female data.

Subsequently, the model was tested for structural invariance by constraining invariant the factor variance/covariance matrix. For this analysis, the noninvariant factor loading parameters were freely estimated across genders (Byrne, Shavelson, & Muthen, 1989). The chi-square change statistic indicated that the additional constraints did not lead to a significant deterioration of fit, $\Delta\chi^2(6, N = 3,374) = 8.04$, $p = ns$.

The last objective of this study was to compare the male and female latent mean structures. This analysis was also

TABLE 6
Summary of Cumulative Tests for Invariance of the Factor Pattern Matrix

Models	χ^2	$\chi^2 df$	$\Delta\chi^2$	Δdf	*p*
1. Baseline	916.66	46	—	—	< .01
2. SOC[a]	925.32	47	8.66	1	< .01
3. AMB	917.35	47	0.69	1	*ns*
4. ANX	919.49	48	2.83	2	*ns*
5. DEP	919.94	49	3.28	3	*ns*
6. SEN	921.62	50	4.96	4	*ns*
7. IMP	921.69	51	5.03	5	*ns*
8. IMP[a,b]	939.17	52	22.51	6	< .01

Note. SOC = Sociability; AMB = Ambitiousness; ANX = Anxiety; DEP = Dependence; SEN = Sensation-Seeking; IMP = Impulsivity.
[a]Noninvariant loadings. [b]This model tests the invariance of the cross-loading of Impulsivity on Neuroticism.

conducted under partial measurement invariance conditions, with the noninvariant loadings and the corresponding intercept terms freely estimated across groups. The constraints on the remaining scale intercepts did not affect the overall fit significantly, $\Delta\chi^2(4, N = 3{,}374) = 7.09$, $p = ns$. The factor intercept for Psychoticism was marginally significant ($z = 1.96$, $p = .05$), indicating higher factor scores for men, whereas those for Extraversion and Neuroticism were not ($z = 0.56$, $p = ns$, and $z = 1.00$, $p = ns$, respectively).

The corrected (Levy, 1967) correlations between the long and short EPP forms were .69, .64, and .87 for Psychoticism, Extraversion, and Neuroticism, respectively (see Table 3 for details). The latter two values were very close to those in Study 1, whereas that for Psychoticism was higher. This can be attributed to the differences in scope between the samples used in the two studies, as Psychoticism is especially susceptible to restriction of range effects in small or relatively homogenous samples (S. B. G. Eysenck et al., 1985).

The factor loadings shown in Table 5, which are very similar to the corresponding within-group completely standardized values, may be used as weights in the estimation of superfactor scores, given that they are based on a very large sample. Nevertheless, this procedure is unlikely to result in substantially improved factor score estimates (Kaiser, 1970; Wainer, 1976).

GENERAL DISCUSSION

The main applied objective of these studies was to develop a revised short form of the EPP. Study 1 indicated that the previous short form (H. J. Eysenck et al., 1996) provides a poor fit to the data, at least when assessed through CFA procedures. In contrast, the revised model derived in Study 1 provides a reasonably good fit. A secondary objective was to examine the invariance of the EPP–SF across men and women. The basic model (see Table 1) was invariant across genders, as were most of the factor loadings and the variance–covariance matrix. As regards the latent means, men scored higher than women on Psychoticism, but not on Extraversion or Neuroticism.

From an applied perspective, the findings of the two studies indicate that the revised EPP–SF can be recommended for the economical assessment of normal adults on the three Eysenckian dimensions. Use of the inventory in applied clinical contexts should be deferred until its psychometric properties and validity have been scrutinized on clinical samples.

Correlations Between the Superfactors

Of particular interest are the correlations between Extraversion and the other two superfactors, which are conceptually orthogonal in the Eysenckian model of personality. More generally, CFA applications have consistently revealed correlations between conceptually independent dimensions both for three-factor as well as for five-factor models (Church & Burke, 1994; Jackson et al., 2000; McCrae, Zonderman, Bond, Costa, & Paunonen, 1996; Moosbrugger & Fischbach, 2002). Although it is possible to reduce these correlations by allowing multiple cross-loadings on nonkeyed factors, it is virtually impossible to eliminate them. Indeed, unconstrained EFAs also have tended to produce substantially correlated factors (e.g., Digman, 1997; Saucier, 2002). The extent to which factors are correlated in EFAs is often masked by the application of orthogonal rotations that assume and impose independence between the dimensions.

The problem is exacerbated when factor scores are estimated, not least because cross-loadings are rarely taken into account in the relevant calculations (Block, 2001). Because the factor correlations in the EPP–SF are unlikely to be CFA artifacts, it should be expected that the corresponding factor scores will exhibit correlations of a similar magnitude. Any application of the questionnaire, then, should heed these correlations and the resulting variance overlap. For example, if both Extraversion and Neuroticism are found to predict a criterion of interest, it will be necessary to examine the relevant partial correlations before concluding that the effects are independent of each other.

A relevant point to consider is that the EPP factors in the long form of the inventory are indeed orthogonal (H. J. Eysenck et al., 1992; Petrides & Furnham, 2001). There are two reasons why this might be so. First, as noted previously, EFA allows part of the covariance between the factors to spill into the factor pattern matrix in the form of cross-loadings. Second, correlations between factors can be spuriously deflated by incorporating factorially complex scales in the questionnaire. The main dimensions in hierarchical models of personality represent points of reference for mapping theoretically continuous factor space and many primary constructs relate to more than one basic dimension (Paunonen, 1998; Saucier, 1992). These relationships have been explicitly acknowledged and modeled in circumplex models of personality (Hofstee, De Raad, & Goldberg, 1992; Wiggins & Trobst, 1999). Nevertheless, incorporating such scales in hierarchical trait structures complicates measurement and can seriously affect the correlations between the factors.

Note that the observed correlations between the superfactors are likely to be inflated by common method variance. Multimethod assessment and subsequent modeling of more than one rating source (e.g., self-report, observer ratings) has been shown to reduce factor correlations and to improve model fit (Piedmont, 1994, 1998).

Cross-Loadings and Invariance

In addition to distorting factor correlations, the presence of factorially complex scales leads to cross-loadings that cannot always be explained theoretically. An interesting cross-loading in the revised EPP–SF concerns the Impulsivity scale. Impulsivity is a typical example of a factorially complex pri-

mary construct that can itself be divided into subcomponents, each having a different pattern of relationships with the three superfactors (S. B. G. Eysenck & Eysenck, 1977; Revelle, 1997). This scale's loadings on Neuroticism and Psychoticism were both modeled in the revised EPP–SF because they are theoretically meaningful, given the pathological aspects of impulsivity (S. B. G. Eysenck & Eysenck, 1977). Consistent with this view, Study 2 showed that EPP Impulsivity correlated much higher with Dickman's (1990) dysfunctional impulsivity, which has negative behavioral consequences, than with functional impulsivity, which may often benefit the individual. Although this difference was based on a small sample, it was statistically significant.

The fact that the cross-loading of Impulsivity is meaningful does not render the scale an ideal marker for Psychoticism. Nevertheless, Impulsivity is one of the better Psychoticism markers in the EPP. As H. J. Eysenck et al. (1992) acknowledged, further work is necessary to replace irrelevant scales and achieve robust measurement of the dimension of Psychoticism in the EPP. Nevertheless, the problem is less serious in the revised EPP–SF because scales with low communalities have been dropped from the inventory.

Although the removal of scales with low communalities will be conducive to attaining replicable simple structures, it may also result in loss of explanatory variance. In other words, it is possible that scales that do not fit in a particular factor space may have considerable validity. Indeed, even scales that clearly belong in a personality factor can sometimes contribute to the prediction and understanding of behavior over and above the factor itself (Paunonen & Ashton, 2001). Moreover, some of the scales that seem to lie outside the Psychoticism–Extraversion–Neuroticism system (e.g., Practical) may well load on the dimensions of personality taxonomies that are broader than the Eysenckian (Costa & McCrae, 1995).

To our knowledge, this is the first study systematically to examine issues of gender invariance in the EPP. With the exception of the loading of Sociability and the cross-loading of Impulsivity, all other loadings were found to be invariant. These results indicate that men and women tend to interpret Sociability and Impulsivity items somewhat differently. With respect to the structural properties of the EPP–SF, the variances and covariances of the superfactors were invariant across genders. Thus, in both solutions, Extraversion was correlated with both Neuroticism and Psychoticism.

Latent Means

In line with results based on various Extraversion scales (Costa & McCrae, 1992b; Costa, Terracciano, & McCrae, 2001; H. J. Eysenck et al., 1992), there were no gender differences in latent means. Also in accord with previous research (S. B. G. Eysenck et al., 1985; H. J. Eysenck et al., 1992; Lynn & Martin, 1997), men scored higher than women on Psychoticism. In contrast, the frequently observed difference in Neuroticism on which women tend to score higher than men (Costa et al., 2001; Lynn & Martin, 1997) was not observed on the large sample of Study 2.

The lack of strong gender differences on a sample as large as that of Study 2 raises an interesting question regarding the magnitude and, more important, the origin of the effects observed in some studies. First, because effect sizes are not always reported in the literature, it is possible to lose sight of the fact that theoretically trivial differences can become "statistically significant" as a direct result of large sample sizes. There were numerous highly significant gender differences in Study 2 (see Table 2). However, with the exception of Risk-Taking, Sensation-Seeking, and Anxious, the effects sizes were very small.

More intriguing is the possibility that even those differences are not differences in the latent variables per se but in the uniquenesses of the indicators. In other words, it is possible that there are gender differences in linear combinations of observed variables, which include error and specific variances, but not in the estimated latent means, which do not. At least with respect to Neuroticism, the results of Study 2 suggest that gender differences in the factor may not be taken for granted. There are certain facets, like Anxiety, on which women tend to score substantially higher than men, but this is not the case for all Neuroticism facets. Thus, depending on the combination of facets that a particular measure of Neuroticism encompasses, gender differences may be accentuated or attenuated. Future research on this issue should focus not only on different personality measures but also on different samples and contexts. The possibility that gender differences in personality dimensions, especially Neuroticism, are minimal is well worth considering in the future both because such differences may have social, cultural, and political implications (Eagly, 1995) and also because they are often taken for granted or invoked to explain empirical findings.

Relationship Between the Long and Short EPP Forms

The correlations between the factors in the long and short forms of the EPP were lower than ideal, particularly as regards Psychoticism and Extraversion. Several of the facets that these two dimensions comprise in the long EPP either did not fit the Psychoticism–Extraversion–Neuroticism scheme at all or fell into complex factor space defined by more than one dimension. In contrast, Neuroticism, which is the most robustly operationalized dimension in the EPP, showed the greatest consistency across the two forms of the inventory.

The exclusion of irrelevant and factorially complex facets from the EPP had some impact on the constitution of the factors. However, it should be clear that the factors in the revised short form are more in line with the theoretical nature of the Eysenckian dimensions than their counterparts in the long form. In other words, dropping 12, mostly substandard,

scales from the long EPP has actually improved the quality of measurement because the information that has been lost was, to a greater or lesser extent, irrelevant to the conceptualization of the Eysenckian dimensions.

CFA and Personality Research

The use of CFA in investigations of personality hierarchies is subject to debate (e.g., Borkenau & Ostendorf, 1990; McCrae et al., 1996; Vassend & Skondral, 1997). The requirement of precise a priori models establishes a stringent starting point for CFA applications. Nevertheless, some established and extensively researched personality inventories meet this preliminary criterion, yet they still fail to fit the data when model fit is evaluated through CFA indexes.

Several reasons are implicated in the rejection of established measures of personality hierarchies by CFA. Before researchers even have a chance to test a model against the stringent CFA indexes, their data must satisfy assumptions that are extremely difficult to be met by comprehensive personality inventories (e.g., multivariate normality; see Raykov, 1998). The violation of assumptions has an impact on the resultant fit indexes and models that are overall acceptable are rejected as inadequate. In addition, at least as far as personality hierarchies are concerned, CFA applications are unnecessarily demanding as regards the precise and complete specification of full factor matrices. In the case of comprehensive personality inventories, this involves a priori determining hundreds of loadings, most of which are theoretically uninteresting. For reasons such as the foregoing, rigid applications of CFA in personality research may not always be relied on to yield valid conclusions and otherwise robust models may not be rejected on the basis of CFA indexes.

Given these limitations, how does CFA inform substantive research in personality? The answer is that CFA forces researchers to reexamine fundamentally important issues that had faded into the background because of the uncritical use of EFA. It is a positive development that issues such as cross-loadings and simple structure have sprung back into the forefront. Although replicable models should not be rejected because of theoretically trivial, nonkeyed loadings, attention should be drawn to large, consistent, and theoretically meaningful cross-loadings. Paradoxically, these are the loadings that are hardest to defend from a measurement perspective because if the interest is in obtaining markers for the major axes of personality, it is difficult to justify the incorporation of variables that fall into complex factor space. Such variables share variance with more than one factor and tend to distort factor correlations.

In revising the EPP–SF through CFA, we attempted to exclude cross-loading scales and to refrain from overfitting the model by freeing parameters that are either trivial or difficult to explain theoretically. The resultant model led to an average fit, at least as judged by Hu and Bentler's (1999) stringent cutoff criteria. However, the statistical properties of fit indexes are not well understood and continue to be studied (e.g., Ogasawara, 2001). Consequently, it remains difficult to appreciate the extent of model misspecification in CFAs. Specifically with respect to the revised EPP–SF, a notable improvement in the fit indexes can be achieved either by overfitting or through a revision of the questionnaire that will require work at the item level, including the development of new markers.

Applications and Validity of the Revised EPP–SF

Before undertaking a revision of the inventory, it will be more important, from a substantive perspective, to investigate the validity of the revised EPP–SF factors. Indeed, the construct validity of psychological concepts rests primarily on the robustness of their nomological networks (i.e., their theoretically supported associations with other constructs and behaviors). The revised EPP–SF has been created largely based on statistical optimization criteria. It is therefore important to establish whether the increased measurement precision achieved is reflected in the predictive and explanatory utility of the inventory. Advantages such as economical assessment, increased common variance, improved simple structure, and enhanced understanding of the gender-specific properties of the inventory are likely to effect stronger, more precise, and more meaningful associations with appropriate criterion variables.

Eysenckian theory (e.g., H. J. Eysenck, 1970) argues for a dimensional representation of mental illness, with normal individuals primarily differing in quantitative, as opposed to qualitative, ways from neurotic and psychotic patients. Clinical assessment on the EPP–SF scales is likely to be useful in the understanding of certain psychopathological conditions, not least because it will be possible to link and interpret the resultant findings with reference to the extensive body of knowledge that the Eysenckian personality model has generated.

CONCLUSIONS

The revised short form of the EPP constitutes an efficient instrument to measure normal adult personality on nine primary facets marking the three Eysenckian dimensions of Psychoticism, Extraversion, and Neuroticism. The structure and factor pattern of the questionnaire are largely invariant across men and women. The former tend to score higher on Psychoticism, but there seem to be no gender differences in either Extraversion or Neuroticism. There are correlations of considerable magnitude between Extraversion and the other two superfactors.

In this article, we have shown that many facets in the long form of the EPP are suboptimal as markers of the Eysenckian dimensions. Although in certain cases CFA can be an overly restrictive tool for assessing personality inventories, it is undoubtedly useful for identifying irrelevant or poorly

operationalized scales. The presence of such scales compromises measurement and often distorts the true relationships between the factors in an inventory. Most such scales were excluded from the revised EPP–SF, which can serve both as a better quality alternative to the long form of the EPP and as the basis for its revision.

ACKNOWLEDGMENTS

This research was supported by an Economic and Social Research Council (ESRC) Postdoctoral Fellowship to K. V. Petrides. Parts of this article were presented at the 23rd Biennial Conference of the Society for Multivariate Analysis in the Behavioral Sciences (Tilburg University, The Netherlands). The long and short forms of the EPP are available from Chris J. Jackson, School of Psychology, University of Queensland, Brisbane, QLD 4072, Australia. E-mail: chrisj@psy.uq.edu.au

REFERENCES

Barrett, P. T., Petrides, K. V., Eysenck, S. B. G., & Eysenck, H. J. (1998). The Eysenck Personality Questionnaire: An examination of the factorial similarity of P, E, N, and L across 34 countries. *Personality and Individual Differences, 25,* 805–819.

Bienvenu, O. J., Nestadt, G., Samuels, J. F., Costa, P. T., Jr., Howard, W. T., & Eaton, W. W. (2001). Phobic, panic, and major depressive disorders and the Five-factor model of personality. *Journal of Nervous and Mental Disease,189,* 154–161.

Block, J. (2001). Millennial contrarianism: The five-factor approach to personality description 5 years later. *Journal of Research in Personality, 35,* 98–107.

Borkenau, P., & Ostendorf, F. (1990). Comparing exploratory and confirmatory factor analysis: A study on the 5-factor model of personality. *Personality and Individual Differences, 11,* 515–524.

Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105,* 456–466.

Cattell, R. B. (1973). *Personality and mood by questionnaire.* San Francisco: Jossey-Bass.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the 16 Personality Factor questionnaire.* Champaign, IL: Institute for Personality and Ability Testing.

Church, A. T., & Burke, P. J. (1994). Exploratory and confirmatory tests of the Big Five and Tellegen's three- and four-dimensional models. *Journal of Personality and Social Psychology, 66,* 93–114.

Costa, P. T., Jr., & McCrae, R. R. (1992a). Four ways five factors are basic. *Personality and Individual Differences, 13,* 653–665.

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

Costa, P. T., Jr., & McCrae, R. R. (1995). Primary traits of the Eysenck P–E–N system: Three and five-factor solutions. *Journal of Personality and Social Psychology, 69,* 308–317.

Costa, P. T., Jr., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology, 81,* 322–331.

Dickman, S. J. (1990). Functional and dysfunctional impulsivity: Personality and cognitive correlates. *Journal of Personality and Social Psychology, 58,* 95–102.

Digman, J. M. (1997). Higher-order factors of the big five. *Journal of Personality and Social Psychology, 73,* 1246–1256.

Eagly, A. H. (1995). The science and politics of comparing women and men. *American Psychologist, 50,* 145–158.

Eysenck, H. J. (1947). *Dimensions of personality.* New York: Praeger.

Eysenck, H. J. (1970). A dimensional system of psychodiagnostics. In A. R. Mahrer (Ed.), *New approaches to personality classification* (pp. 169–208). New York: Columbia University Press.

Eysenck, H. J. (1990). Genetic and environmental contributions to individual-differences: The three major dimensions of personality. *Journal of Personality, 58,* 245–261.

Eysenck, H. J. (1992). Four ways five factors are not basic. *Personality and Individual Differences, 13,* 667–673.

Eysenck, H. J. (1997). Personality and experimental psychology: The unification of psychology and the possibility of a paradigm. *Journal of Personality and Social Psychology, 73,* 1224–1237.

Eysenck, H. J., Barrett, P. T., Wilson, G. D., & Jackson, C. J. (1992). Primary trait measurement of the 21 components of the P–E–N system. *European Journal of Psychological Assessment, 8,* 109–117.

Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach.* New York: Plenum.

Eysenck, H. J., & Wilson, G. D. (1991). *The Eysenck Personality Profiler.* London: Corporate Assessment Network.

Eysenck, H. J., Wilson, G. D., & Jackson, C. J. (1996). *The Eysenck Personality Profiler (short).* Guildford: Psi-Press.

Eysenck, S. B. G., & Eysenck, H. J. (1977). The place of impulsiveness in a dimensional system of personality description. *British Journal of Social and Clinical Psychology, 16,* 57–68.

Eysenck, S. B. G., Eysenck, H. J., & Barrett, P. T. (1985). A revised version of the Psychoticism scale. *Personality and Individual Differences, 6,* 21–29.

Francis, L. J., Jones, S. H., Jackson, C. J., & Robbins, M. (2001). The feminine personality profile of male Anglican clergy in Britain and Ireland: A study employing the Eysenck Personality Profiler. *Review of Religious Research, 43,* 14–23.

Furnham, A., Petrides, K. V., Jackson, C. J., & Cotter, T. (2002). Do personality factors predict job satisfaction? *Personality and Individual Differences, 33,* 1325–1342.

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

Hofstee, W. K. B., De Raad, B., & Goldberg, L. R. (1992). Integration of the Big Five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology, 63,* 146–163.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Jackson, C. J., Furnham, A., Forde, L., & Cotter, T. (2000). The structure of the Eysenck Personality Profiler. *British Journal of Psychology, 91,* 223–239.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Chicago: Scientific Software.

Kaiser, H. F. (1970). A second generation Little Jiffy. *Psychometrika, 35,* 401–417.

Levy, P. (1967). The correction for spurious correlation in the evaluation of short-form tests. *Journal of Clinical Psychology, 23,* 84–86.

Lynn, R., & Martin, T. (1997). Gender differences in extraversion, neuroticism, and psychoticism in 37 nations. *Journal of Social Psychology, 137,* 369–373.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51,* 201–226.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111,* 490–504.

McCrae, R. R., & Costa, P. T., Jr. (1999). A five-factor theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 139–153). New York: Guildford.

McCrae, R. R., Zonderman, A. B., Bond, M. H., Costa, P. T., Jr., & Paunonen, S. V. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus procrustes rotation. *Journal of Personality and Social Psychology, 70,* 552–566.

Moosbrugger, H., & Fischbach, A. (2002). Evaluating the dimensionality of the Eysenck Personality Profiler–German version (EPP–D): A contribution to the super three vs. big five discussion. *Personality and Individual Differences, 33,* 191–211.

Muris, P., Schmidt, H., Merckelbach, H., & Rassin, E. (2000). Reliability, factor structure and validity of the Dutch Eysenck Personality Profiler. *Personality and Individual Differences, 29,* 857–868.

Ogasawara, H. (2001). Standard errors of fit indices using residuals in structural equation modeling. *Psychometrika, 66,* 421–436.

Paunonen, S. V. (1998). Hierarchical organization of personality and prediction of behavior. *Journal of Personality and Social Psychology, 74,* 538–556.

Paunonen, S. V., & Ashton, M. C. (2001). Big five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology, 81,* 524–539.

Petrides, K. V., & Furnham, A. (2001). Trait emotional intelligence: Psychometric investigation with reference to established trait taxonomies. *European Journal of Personality, 15,* 425–448.

Piedmont, R. L. (1994). Validation of the NEO PI–R observer form for college students: Towards a paradigm for studying personality development. *Assessment, 1,* 259–268.

Piedmont, R. L. (1998). *The revised NEO Personality Inventory: Clinical and research applications.* New York: Plenum.

Raykov, T. (1998). On the use of confirmatory factor analysis in personality research. *Personality and Individual Differences, 24,* 291–293.

Revelle, W. (1997). Extraversion and impulsivity: The lost dimension? In H. Nyborg (Ed.), *The scientific study of human nature: Tribute to Hans J. Eysenck at eighty* (pp. 189–212). New York: Pergamon.

Saucier, G. (1992). Benchmarks: Integrating affective and interpersonal circles with the Big-Five personality factors. *Journal of Personality and Social Psychology, 62,* 1025–1035.

Saucier, G. (2002). Orthogonal markers for orthogonal factors: The case of the big five. *Journal of Research in Personality, 36,* 1–31.

Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement, 54,* 837–847.

Vassend, O., & Skrondal, A. (1997). Validation of the NEO Personality Inventory and the Five-factor model. Can findings from exploratory and confirmatory factor analysis be reconciled? *European Journal of Personality, 11,* 147–166.

Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin, 83,* 213–217.

Wiggins, J. S., & Trobst, K. K. (1999). The fields of interpersonal behavior. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 653–670). New York: Guildford.

Wilson, G. D., & Jackson, C. (1994). The personality of physicists. *Personality and Individual Differences, 16,* 187–189.

K. V. Petrides
Institute of Education
University of London
25 Woburn Square
London WC1H 0AA, England
E-mail: k.petrides@ioe.ac.uk

Received June 24, 2002
Revised May 27, 2003