

Consensual Validity Parameters of the Zuckerman–Kuhlman Personality Questionnaire: Evidence From Self-Reports and Spouse Reports

Montserrat Gomà-i-Freixanet and Andreas A. J. Wismeijer

*Department of Health Psychology
Autonomous University of Barcelona*

Sergi Valero

*Department of Psychiatry
Hospital Universitari de la Vall d'Hebron
Barcelona, Catalonia, Spain*

One way of validating questionnaire responses is correlating them with ratings made by external assessors who know the ratee well: This is known as consensual validity. In this study, we assessed the consensual validity of the Zuckerman–Kuhlman Personality Questionnaire (ZKPQ; Zuckerman, 2002; Zuckerman, Kuhlman, Joireman, Teta, & Kraft, 1993). A multitrait-multimethod matrix of self-reported and spouse reported personality ratings was used to establish convergent and discriminant validity by means of Campbell & Fiske's (1959) evaluative criteria. Self-reports of 86 men and 85 women were correlated with their spouses' reports. Intraclass correlations ranged from .47 to .63 for the 5 dimensions, providing strong evidence of convergent and discriminant validity. The results obtained favor the use of the ZKPQ as a valid self-report measure of personality traits.

The alternative Five-factor model (AFFM) proposed and described by Zuckerman, Kuhlman, Thornquist, and Kiers (1991) originated as an alternative to the Five-factor model (FFM; Digman, 1990) to make up for the latter's supposed lack of explanatory power. The origins of the FFM and the AFFM were quite different. The former originated in research on the lexical properties of adjectives in the language pertaining to personality descriptors. Costa, McCrae, and Arenberg (1980) started with a three-factor model (Extraversion, Neuroticism, and Openness to Experience), resembling Eysenck's (1967) model at least in the first two factors. Costa and McCrae (1985) later added two more factors (Agreeableness and Conscientiousness) to bring the model closer to the five factors identified in lexical analyses. The model is essentially a descriptive one designed to tap those dimensions of personality identified in the lexical analyses of dictionaries and factor analyses of reduced subsets of the personality relevant words. It is essentially atheoretical in its origin and descriptive of traits that are uniquely human and do not translate easily into compara-

tive descriptions of animal behavior traits (e.g., Conscientiousness, Agreeableness).

The AFFM was being developed at the same time as Costa and McCrae (1985) were developing their own model. In preparation for his book on the *Psychobiology of Personality*, Zuckerman (1991) began looking for a framework to describe personality traits with biological-evolutionary roots and a potential for comparative analyses. Zuckerman et al. (1991) used questionnaire scales that had already been widely used in human psychobiological research and/or in studies of temperament in children and adults. Factor analyses using several markers for each hypothesized trait consistently yielded five factors reliably identifiable across genders (Zuckerman, Kuhlman, & Camac, 1988; Zuckerman et al., 1991). Both Eysenck (1992a, 1992b) and Zuckerman (1992) pointed out that one way of determining which are the basic traits of personality is using a nomological network or a theoretical underpinning to guide taxonomic studies and lead to theoretical advances and that psychobiological studies of personality provide data for the understanding of the

neurobiological and genetic underpinnings of personality. Relying only on the encoding of personality traits in language is treacherous, as this encoding probably reflects the observability of these traits in social interactions and may not necessarily mirror the proportional biological relevance of the traits (Zuckerman, 1992; Zuckerman, Kuhlman, Joireman, Teta, & Kraft, 1993). Therefore, using psychobiological data instead allows researchers to explore the biological origins of personality (Eysenck, 1992a).

The Zuckerman–Kuhlman Personality Questionnaire (ZKPQ; Zuckerman, 2002; Zuckerman et al., 1993) was developed to measure the dimensions that constitute the AFFM, namely, Neuroticism-Anxiety (N-Anx), Activity (Act), Sociability (Sy), Impulsive Sensation-Seeking (ImpSS), and Aggression-Hostility (Agg-Host). Since its first publication in 1993, the ZKPQ has undergone extensive psychometric testing, targeting both reliability and validity parameters. Concerning criterion validity, the ZKPQ has been tested in different populations such as team sportsmen and sportswomen (O'Sullivan, Zuckerman, & Kraft, 1998), prostitutes (O'Sullivan, Zuckerman, & Kraft, 1996), and cocaine abusers (Ball, 1995). Several cross-cultural studies have been performed as well using translations in Germany (Ostendorf & Angleitner, 1994), Japan (Shiomi et al., 1996), Israel (I. Montag, 2001), China (Wu et al., 2000), Italy (De Pascalis & Russo, 2003), and Spain (Gomà-i-Freixanet, Valero, Puntí, & Zuckerman, 2004; Kuhlman, Zuckerman, Gomà-i-Freixanet, & Shiomi, 2003).

Our aim of this study was to provide data on the consensual validity of the ZKPQ using the Catalan translation by Gomà-i-Freixanet et al. (2004). The psychometric evaluation of this version showed good internal consistency and discriminant validity of scales. The mean Cronbach's alpha for the ZKPQ scales was .76, with values ranging from .67 to .84, indicating adequate internal consistency within scales. Also, the correlations among scales ranged from $-.19$ to $.22$, with an absolute mean interscale correlation of $.06$, indicating adequate discriminant validity. Furthermore, gender differences in means were in accordance with the original U.S. sample, and the original U.S. factor structure was largely replicated with Tucker's congruence coefficients generally being in the 90s (Gomà-i-Freixanet et al., 2004).

The construct validity of a test is defined by the degree to which the test measures the variable(s) that it is designed to measure. This is easy to establish when one has a criterion against which the test can be evaluated, but unfortunately, this is not always the case. One way of overcoming this problem could be what is known as "consensual validation" (McCrae, 1982), meaning that one can validate questionnaire responses by correlating them with ratings made by external assessors who know the ratee well. This method thus provides both reports from observers who can interpret specific overt behaviors as evidence of underlying traits and self-reports that provide unique access to the private thoughts of the rated individual. One of the advantages of this

approach (McCrae & Costa, 1983) is that the artifacts that may influence self-reports (e.g., acquiescence, social desirability, and defensive responding) in general may be considered to be independent from those that influence observer reports (e.g., halo effects or effects of stereotypes). An agreement between self-reports and observer reports therefore constitutes powerful evidence of consensual validation. In a former article, Gomà-i-Freixanet (1997) gave a detailed overview of factors affecting the accuracy or agreement between self-reports and observer reports. Some of the most important conditions leading to improved accuracy of reports are the instrument itself and the acquaintance between the target and the rater. McCrae & Costa (1983) stated that one could increase the correspondence between the two sources by employing the same instrument for both the target individual and the rater. Regarding the target–rater acquaintance, it is known that self–peer convergence and interrater reliability increase with the degree of acquaintance between the judge and the target. Norman and Goldberg (1966) pointed out that, judged against the external criteria of self-reports, raters who have had longer acquaintances with the targets give consistently more accurate ratings. Several authors (e.g., McCrae, 1982; Watson & Clark, 1991) found that spouses are more accurate in assessing personality traits than friends, and friends are more accurate than strangers.

This research is the first that aimed to assess the validity of the ZKPQ by means of the consensual validation method using the multitrait-multimethod matrix (Campbell & Fiske, 1959). This approach allows us to assess the convergent and discriminant validity of the ZKPQ scales. By convergent validity, it is meant that measures of the same trait should be in agreement even if they are measured with different methods, and discriminant validity refers to different traits that should be distinguished from one another even if they are measured with the same method. As one can see from the previous reported literature, self–peer agreement increases with the degree of acquaintance between the rater and the ratee. Therefore, we used peer raters well acquainted with their targets: spouses. Thus, the same questionnaire was answered by both members of the couple, spouses were well acquainted with their targets, and using both self-reports and observer reports (from now on referred to as "spouse reports"), we controlled for the artifacts affecting reports in general.

METHOD

Participants

Participants were friends, parents, and relatives of undergraduate and postgraduate students they recruited to participate in the study. Also, some participants were recruited by one of the authors himself (A. Wismeijer), although they did not know the author. The primary convenience sample consisted of 88 couples. As occasionally some couples returned

self-reports or spouse reports for one person only or because reports were not answered properly (defined by us as having $\geq 10\%$ of double or missing answers), data of 86 men and 85 women were eligible for analysis. Age ranged from 19 to 75 years ($M = 38.21$, $SD = 13.14$). Mean amount of years the couples are living together was 13.33 years, with a range of .08 to 44 years ($SD = 11.66$). The mode of level of education was high school and the socioeconomic status was middle class. All respondents participated without inducements and voluntarily in the study.

Materials

Data were obtained using the Catalan translation of the ZKPQ (Gomà-i-Freixanet et al., 2004). The ZKPQ consists of 99 dichotomous items (in sentence format and true–false response set) covering five scales and an additional Infrequency (Infreq; 10 items) scale that allows eliminating participants with careless responding. The dimensions measured are N-Anx (19 items), Act (17 items), Sy (17 items), ImpSS (19 items), and Agg-Host (17 items).

Procedure

The study took place in the province of Barcelona, Catalonia, Spain. All participants spoke Catalan as their primary language. Catalan is one of the four official languages in Spain spoken by more than 7 million inhabitants in different countries (Andorra, France, and Italy). Each individual participant was provided with an envelope containing an introductory letter, the ZKPQ, written instructions, and two answering forms, one for himself or herself and the other to be answered as spouse (i.e., the participants were instructed to answer the ZKPQ as how they see themselves and how they see their spouse). The introductory letter explained globally the goals of the study (“the study you will collaborate in attempts to evaluate the functioning of the Catalan translation of an American questionnaire”), with no reference being made to the consensual agreement component of the study. The letter was signed by the principal investigators (M. Gomà-i-Freixanet and A. Wismeijer). The written instructions stressed to the participants to answer the questionnaire alone and without help of the spouse and instructed the participant to put the answered forms after completion in the provided envelope and seal it immediately. The instructions (written in Catalan) followed the following format:

- Please answer the questionnaire when you are **alone and without** help of your partner.
- First fill out the questionnaire as how you see yourself and **only then** fill out again the questionnaire as how you see your partner (i.e., how you see your partner and **NOT** how you think your partner will fill out the questionnaire).
- It is very important to answer the questionnaire honestly.

- When you filled out both answering forms, please put the **questionnaire and the answering forms** in the envelope and close it to guarantee your privacy.

Each envelope was precoded as well as the answering forms; therefore, all questionnaires were answered anonymously and confidentiality was guaranteed. One member of the couple returned the sealed envelopes to the investigator or the student that recruited the couple, in which case the student subsequently returned the envelopes to the investigators. No envelope was returned opened or with the seal damaged. Thus, from each couple, two sealed envelopes were returned containing in total four answering forms. Apart from the written instructions each participant received, students were first well instructed on how to give the instructions to fill in the questionnaires before recruiting their family members and/or friends so they could repeat the instructions if any doubts would arise after reading the written instructions. Telephone numbers of the principal authors (M. Gomà-i-Freixanet and A. Wismeijer) were included in the letter as well to resolve any doubts. No participant contacted us for questions or doubts. As the study was not intrusive in any sort, neither informed consent waivers nor participant debriefing following participation were required.

Analyses

A multitrait-multimethod matrix was used to determine the consensual validity by comparing the self-reports and spouse reports and establishing the degree of agreement on the different scales of the questionnaire. This approach is based on the assumption that if our measures are valid, we should expect to see a certain pattern emerge among the correlations in the multitrait-multimethod matrix. Specifically, correlations in the principal validity diagonal should be larger than the ones in the same row and column. The validity diagonal values show the correlation of the same trait across different methods. In this way, we could determine whether self-ratings differed from ratings the participants received by their spouses. Therefore, intraclass correlations (ICCs) between self-reports and spouse reports were computed for each scale, for the total sample as well as for both genders to obtain additional validity data and to ascertain if there was any differential pattern in both genders. Pearson correlation measures the intensity of the linear association between two variables but does not give information on the observed agreement, thus ignoring differences in rater’s levels of response. Contrarily, the ICC coefficient is a more accurate statistic than Pearson’s zero order correlation, as the ICC is sensitive to magnitudinal differences between the variables (Bland & Altman, 1986; Prieto, Lamarca, & Casado, 1998). Using traditional zero order correlations, a high correlation between self-reports and spouse reports on a given scale might be found, even though both raters rate each other at a different level. As long as this difference is consistent, it will

not lower the Pearson's correlation coefficient. The ICC solves this problem by taking the differences in magnitude between the scores into account. Using this technique, high yet consistent differences between variables will therefore not lead to an inflated correlation.

The Type I error rate per comparison was set by default to .05. As the ZKPQ has a total of six scales, and age was included in the analysis, a Bonferroni correction for multiple tests of $.05/7 = .007$ was used for the between-participant unpaired *t* tests. A correction of $.05/5 = .01$ was used for the self-reports and spouse reports paired *t* tests, as age and Infreq were not included in these analyses. Finally, an additional principal components analysis (PCA) of the 10 ZKPQ scales scores of self-ratings and spouse ratings was executed. We used this statistical technique, as it fitted optimally to the number of actual observations. Our sample size prevented us from conducting confirmatory factor analysis via structural equation modeling. We forced a five-factor solution and subsequently used the Varimax method for rotation. We predicted that if the five factors of the ZKPQ (Infreq scale was left out) would genuinely show consensual validity, each factor in the rotated matrix would be heavily loaded by both the self-ratings and spouse ratings of the same scale.

RESULTS

Descriptive Statistics

Descriptive statistics from the sample are shown in Table 1. Men and women did not differ significantly on age, although the former were on average 3.20 years older. Means of the scales of the ZKPQ are very similar to the original U.S. version and to the Catalan version (Gomà-i-Freixanet et al., 2004), and they follow the general trend found in men and women, that is, women score higher on N-Anx and Sy and lower on Imp-SS, and Infreq, although these differences only reached significance on the N-Anx scale.

In examining relations between self-data and spouse data, we compared the means across the two methods of data attainment. As was discussed earlier (Gomà-i-Freixanet, 1997), two different patterns might be expected on theoretical grounds. First, researchers who emphasize the biased nature of self-report data would predict that self-raters will respond in a more socially desirable manner (i.e., higher levels on Sy and generally lower levels on N-Anx, ImpSS, Agg-Host and Infreq) than their spouses. Conversely, those who emphasize the biased nature of spouse-report data would predict that spouses will assign generally lower levels of neuroticism than the self-raters because this trait is less externally observable (Johnson, 1997). Table 2 presents the means and standard deviations for the ZKPQ dimensions for self-reports and spouse reports. The data did not support either contention, as we did not find any significant difference between self-reports and spouse reports.

TABLE 1
Means and Standard Deviations
for Self-Reported ZKPQ Scales for Men
and Women and *t* Test Comparisons

Variable	Men ^a		Women ^b		<i>t</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Age (years)	39.84	13.50	36.64	12.67	1.57	.25
ZKPQ						
N-Anx	5.92	4.06	9.84	4.88	-5.73*	-.88
Act	8.35	3.55	7.56	3.79	1.40	.22
Sy	6.58	3.63	6.89	3.51	-0.59	-.09
ImpSS	7.58	4.09	7.08	4.16	0.79	.12
Agg-Host	7.12	3.33	6.41	2.79	1.50	.23
Infreq	2.13	1.57	1.66	1.48	2.01	.31

Note. ZKPQ = Zuckerman-Kuhlman Personality Questionnaire; N-Anx = Neuroticism-Anxiety; Act = Activity; Sy = Sociability; ImpSS = Impulsive Sensation Seeking; Agg-Host = Aggression-Hostility; Infreq = Infrequency.

^a*n* = 86. ^b*n* = 85.

**p* < .001, two-tailed.

TABLE 2
Means and Standard Deviations for ZKPQ
Scales for Self-Reports and Spouse Reports
and *t* Test Comparisons With Paired Data

Scale	Self-Reports ^a		Spouse Reports ^b		<i>t</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
ZKPQ						
N-Anx	7.83	4.90	7.82	4.80	0.04	.00
Act	7.92	3.70	7.95	3.72	-0.10	-.01
Sy	6.77	3.55	7.15	4.14	-1.34	-.07
ImpSS	7.37	4.11	7.29	4.19	0.28	.02
Agg-Host	6.76	3.06	6.74	3.75	0.07	.00

Note. ZKPQ = Zuckerman-Kuhlman Personality Questionnaire; N-Anx = Neuroticism-Anxiety; Act = Activity; Sy = Sociability; ImpSS = Impulsive Sensation Seeking; Agg-Host = Aggression-Hostility.

^a*n* = 171. ^b*n* = 171.

To examine the internal consistency of self-reports and spouse reports, we calculated Cronbach's alpha for both. The mean self-reported α for the ZKPQ scales was .78, with values ranging from .69 to .87. Table 3 shows that N-Anx has the highest internal consistency and Agg-Host the lowest. These coefficients are adequate and very similar to those found in the original U.S. version and even slightly higher than those of the Catalan version. The mean spouse-reported α for the same scales was .81, with values ranging from .77 to .86.

Heteromethod Correlations

Subsequently, ICCs were computed to determine the convergent and discriminant validity of the ZKPQ scales. Table 4 shows the multitrait-multimethod matrix with the heteromethod correlations between the self-reports and spouse reports. The principal diagonal of this table, underlined, con-

TABLE 3
Internal Consistency for Self-Reported and Spouse Rated ZKPQ Scales

Scale	Alpha Coefficients	
	Self-Reports	Spouse Reports
ZKPQ		
N-Anx	.87	.86
Act	.76	.77
Sy	.78	.84
ImpSS	.80	.80
Agg-Host	.69	.79

Note. ZKPQ = Zuckerman-Kuhlman Personality Questionnaire; N-Anx = Neuroticism-Anxiety; Act = Activity; Sy = Sociability; ImpSS = Impulsive Sensation-Seeking; Agg-Host = Aggression-Hostility.

TABLE 4
Multitrait-Multimethod Matrix for Self-Reports and Spouse Reports

Scale	Spouse Reports				
	N-Anx	Act	Sy	ImpSS	Agg-Host
Self-Reports					
N-Anx	<u>.63***</u>	-.11	-.07	.04	.16*
Act	-.02	<u>.47***</u>	.24**	.28***	.16*
Sy	-.15	.16*	<u>.54***</u>	.34***	.20**
ImpSS	-.01	.10	.34***	<u>.63***</u>	.22**
Agg-Host	.00	-.14	.08	.10	<u>.53***</u>

Note. Convergent correlations are underlined. N-Anx = Neuroticism-Anxiety; Act = Activity; Sy = Sociability; ImpSS = Impulsive Sensation-Seeking; Agg-Host = Aggression-Hostility.

* $p < .05$, two-tailed. ** $p < .01$, two-tailed. *** $p < .001$, two-tailed.

tains the convergent validity coefficients for the five scales. Overall, the results indicate a clear convergent and discriminant pattern. In terms of convergent validity, all scales showed a significant level of self-peer agreement.

Total sample convergent correlations ranged from .47 to .63 (absolute mean interscale correlation = .56). Clearly, all convergent correlations were far higher than the so-called “.3 barrier” (Mischel, 1968). In contrast, all but two of the off-diagonal values assessing discriminant validity stayed below the .3 barrier. Using the criterion that convergent correlations should be higher than any other values in its row or column of the heteromethod matrix (Campbell & Fiske, 1959), we found that all the scales displayed a good level of discriminant validity. These results support the construct validity of these specific scales. Self-reports and spouse reports showed a significant level of agreement on each of the scales.

Table 5 shows convergent correlations between self-reports and spouse reports for the total sample as well as for men and women separately to analyze any gender-based differences in ratings, that is, whether gender can influence the convergence of spouse ratings on any scale. Note that the correlations provided in the first column of Table 5 are the same as the principal diagonal correlations underlined in Table 4. The

95% confidence intervals are provided for these coefficients to show the estimation of population parameters. Convergent correlations ranged from .48 to .60 for men (absolute mean interscale correlation = .54) and .45 to .71 for women (absolute mean interscale correlation = .55).

The results for men and women followed the general pattern found in the total sample; there were, however, differences in absolute values between men and women in the magnitude of correlations in two scales: Sy and ImpSS. Agreement between self-reports and spouse reports was higher for Sy when women assessed men, whereas for the ImpSS scale, agreement was higher when men assessed women.

To evaluate the possible influence of the years the partners were living together on the level of agreement between self-reports and spouse reports, we generated a new variable named *consensus*, which consisted of subtracting for each scale the self-scores from the spouse scores. This new variable gave us a measure of the difference between self-reports and spouse-reports scores. None of the correlations between consensus and years living together, which ranged from 0 to 44 years, were significant for any of the scales: N-Anx ($r = .13$), Act ($r = .04$), Sy ($r = .05$), ImpSS ($r = .04$), Agg-Host ($r = .07$), and Infreq ($r = .09$).

PCA

Finally, Table 6 shows the results of the factor analysis of the self-reported and spouse-reported scales (excluding the Infreq scale) using a PCA followed by normalized varimax rotation of the five scales of the ZKPQ. The rationale behind this methodological strategy was that if the ZKPQ showed a high degree of consensual validity, then the responses made by both assessors should be highly consistent among them for each one of the scales. A PCA that contemplates the assessments of both assessors should consistently gather in each factor self-reports

TABLE 5
Intraclass Correlations With 95% Confidence Interval Between Self-Reports and Spouse Reports for the Total Sample and for Men and Women Separately

Scale	Entire Sample ^a		Men ^b		Women ^c	
	ICC	CI	ICC	CI	ICC	CI
ZKPQ						
N-Anx	<u>.63</u>	.53 to .71	.54	.37 to .67	.59	.44 to .72
Act	.47	.34 to .58	.48	.29 to .62	.46	.28 to .62
Sy	.54	.42 to .64	.60	.45 to .72	.45	.26 to .61
ImpSS	.63	.53 to .71	.56	.39 to .69	.71	.58 to .80
Agg-Host	.53	.41 to .63	.54	.37 to .68	.53	.36 to .67

Note. All the correlations were statistically significant at $p < .001$, two-tailed. ICC = intraclass correlations; CI = 95% confidence interval; ZKPQ = Zuckerman-Kuhlman Personality Questionnaire; N-Anx = Neuroticism-Anxiety; Act = Activity; Sy = Sociability; ImpSS = Impulsive Sensation-Seeking; Agg-Host = Aggression-Hostility.

^a $n = 171$. ^b $n = 86$. ^c $n = 85$.

TABLE 6
Result of the Principal Components
Analysis Followed by a Varimax Rotation
of the Five ZKPQ Scales

Scale	Rotated Factor Loadings				
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
ImpSS self	<u>.89</u>	.04	.21	.14	.08
ImpSS spouse	<u>.81</u>	-.01	.04	.31	.19
N-Anx spouse	-.13	<u>.92</u>	.02	-.04	.06
N-Anx self	.17	<u>.86</u>	.12	-.11	-.17
Agg-Host self	.14	.01	<u>.89</u>	-.01	-.06
Agg-Host spouse	.10	.16	<u>.82</u>	.21	.06
Sy spouse	.22	-.01	.04	<u>.85</u>	.16
Sy self	.18	-.15	.15	<u>.82</u>	.06
Act spouse	.03	-.02	-.19	.16	<u>.85</u>
Act self	.23	-.08	.20	.06	<u>.81</u>
% variance	16.49	16.31	16.21	15.89	14.84

Note. Factorial weights corresponding to pairs of self-ratings and spouse ratings are underlined. ZKPQ = Zuckerman-Kuhlman Personality Questionnaire; N-Anx = Neuroticism-Anxiety; Act = Activity; Sy = Sociability; ImpSS = Impulsive Sensation-Seeking; Agg-Host = Aggression-Hostility.

and spouse reports. It can be seen that the five rotated components are made up of the combinations of the self-reported and the corresponding spouse-reported ratings of the same scale. The five-factor solution together explained 79.74% of the variance. It is noteworthy to mention the magnitude of the factorial weights that corresponded to each pair of self-ratings versus spouse ratings for each scale that were characterized by highly and very similar factorial weights. Furthermore, there was a high discrepancy between these pairs of assessments and the rest of factorial weights in each factor. These results add additional data about the convergent and discriminant validity of ZKPQ scales.

DISCUSSION

The data obtained with this sample provide strong evidence for the satisfying psychometric properties of the Catalan version of the ZKPQ in general and its consensual validity in particular. With respect to the psychometric properties, gender differences in means among the scales were in the same direction and sense of those found in the original U.S. sample (Zuckerman & Kuhlman, 1993), in a different sample of university students (Gomà-i-Freixanet et al., 2004), and in other cross-cultural samples (e.g., Kuhlman et al., 2003). In general, women score higher on N-Anx and Sy, lower on ImpSS and Infreq, and similar to men, on Act and Agg-Host. With regard to the internal consistency, the results are in accordance to those previously found in other studies and slightly higher than those previously found in another sample with university students (Gomà-i-Freixanet et al., 2004).

In relation to the consensual validity parameters, the data obtained provide clear evidence for the consensual validity

of the personality dimensions assessed by the ZKPQ. Convergent correlations well above the .3 barrier were found. These results seem unlikely to stem from the artifacts of social desirability, acquiescence, extreme responding, or shared stereotypes because these sources of variance are generally not found simultaneously in self-reports and observer reports. Furthermore, as Eysenck and Eysenck (1985) pointed out, although such factors are not always entirely absent, they play only a relatively small part in such personality questionnaires, except under special conditions of motivation in which dissimulation may assume a more prominent role. These conditions generally are present in selection processes, forensic contexts, or in the case of not obtaining data anonymously. Neither condition was present in our study. Additional data come from the comparison of the mean scores for self-reports and spouse reports in that they did not show significant differences.

The magnitude of the convergent correlations—generally .5 to .6—was larger than typically has been reported (e.g., Borgatta, 1964; McCrae & Costa, 1987). Two factors could probably have contributed to the relatively higher correlations: the psychometric adequacy of the instrument itself and/or the nature of the raters. With reference to the instrument, these results seem to add data on the sound psychometric properties of the ZKPQ, specifically on validity. With respect to the second factor, it has been well established (e.g., Norman & Golberg, 1966; Watson & Clark, 1991) that the choice of well-qualified raters enhances the accuracy of personality assessments. On the whole, spouses were well acquainted with the participants they rated, most over a period of many years. Our data also address another issue pointed out by McCrae & Costa (1989). When reliable and valid measures are used, the correlations considerably exceed the .3 barrier, being better characterized as facing the “.6 barrier.” It seems likely that the correlations found are near the ceiling for self-other agreement. Raters will always diverge to some extent from the individual’s phenomenological view of himself or herself, and indeed, it would be disconcerting to think others could know us as intimately as we know ourselves.

Another issue that comes from our results and merits some comment is the lack of correlation between the years the partners were living together and the degree of agreement between self-reports and spouse reports. This lack of correlation corroborates some previous studies with the Eysenck Personality Questionnaire (Gomà-i-Freixanet, 1997) and with other instruments (Buss, 1984; Caspi, Herbener, & Ozer, 1992). It has already been mentioned that self-peer convergence increases with the degree of acquaintance between the judge and the target and that spouses are more accurate in assessing personality traits than friends and friends more accurate than strangers. Therefore, it seems as if there is a differential degree of agreement related to the degree of acquaintance, but the degree of agreement does not increase with the years of living together in this sample.

Additional data about convergent and discriminant validity of the ZKPQ scales stem from the PCA executed with self-reports and spouse reports scales scores. Convergent validity indexes are characterized by high and highly similar loadings of the self-reports and spouse reports of the same scale on a given factor. Discriminant validity indexes are characterized by relatively low loadings of the remaining scales on that given factor. The results unambiguously show this pattern of behavior for each pair of reports of the same scale and in every factor.

To summarize, the magnitude of the convergent correlations found in this study among self-reports and spouse reports were larger than typically reported and near the ceiling for self-other agreement (.6 barrier), and the data obtained from the PCA indicate that the consensual validity parameters of the ZKPQ are adequate, thus advocating the use of the self-reported ZKPQ as a valid instrument for personality assessment. Moreover, the results of this study when discussed in the context of previous research undergone in other cultures provide support for the cross-cultural validity of the ZKPQ and the AFFM as a method of evaluating personality functioning.

Finally, the use of self-reports and spouse reports obtained with reliable and valid instruments in personality assessment is not only useful for research purposes but for applied purposes as well. For example, it could be useful in assisting for diagnosis, tailoring treatment techniques, and predicting compliance and success (Mutén, 1991). Comparisons among self-ratings and spouse ratings could also be useful in some psychotherapeutic formulations and in some assessment contexts that clearly encourage biased responding such as in selection settings or in forensic contexts. It would also be a useful alternative in cases in which the ability to report accurately is impaired by cognitive deficits or by psychiatric illness. Thus, using spousal ratings to assess personality in general is a potentially fruitful avenue for future research.

ACKNOWLEDGMENTS

This research was supported by the Autonomous University of Barcelona, Spain with Grant EMER99-18. We thank the students and participants that cooperated in the study and Carmina Puig Sobrevals for her logistical help. Comments by Marvin Zuckerman and by the anonymous reviewers are gratefully acknowledged.

REFERENCES

Ball, S. (1995). The validity of an alternative Five-factor measure of personality in cocaine abusers. *Psychological Assessment*, 7, 148-154.
 Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1, 307-310.
 Borgatta, E. F. (1964). The structure of personality characteristics. *Behavioral Science*, 9, 8-17.

Buss, D. M. (1984). Marital assortment for personality dispositions: Assessment with three different data systems. *Behavior Genetics*, 14, 111-123.
 Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
 Caspi, A., Herbener, E. S., & Ozer, D. J. (1992). Shared experiences and the similarity of personalities: A longitudinal study of married couples. *Journal of Personality and Social Psychology*, 62, 281-291.
 Costa, P. T., Jr., & McCrae, R. R. (1985). *The NEO Personality Inventory*. Odessa, FL: Psychological Assessment Resources.
 Costa, P. T., Jr., McCrae, R. R., & Arenberg, D. (1980). Enduring dispositions in adult males. *Journal of Personality and Social Psychology*, 38, 793-800.
 De Pascalis, V., & Russo, P. (2003). Zuckerman-Kuhlman Personality Questionnaire: Preliminary results of the Italian version. *Psychological Reports*, 92, 965-974.
 Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417-440.
 Eysenck, H. J. (1967). *The biological basis of personality*. Springfield, IL: Thomas.
 Eysenck, H. J. (1992a). Four ways five factors are not basic. *Personality and Individual Differences*, 13, 667-673.
 Eysenck, H. J. (1992b). A reply to Costa and McCrae. P or A and C—The role of theory. *Personality and Individual Differences*, 13, 867-868.
 Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences*. London: Plenum.
 Gomà-i-Freixanet, M. (1997). Consensual validity of the EPQ: Self-Reports and Spouse-Reports. *European Journal of Psychological Assessment*, 13, 179-185.
 Gomà-i-Freixanet, M., Valero, S., Puntí, J., & Zuckerman, M. (2004). Psychometric properties of the Zuckerman-Kuhlman Personality Questionnaire in a Spanish sample. *European Journal of Psychological Assessment*, 20, 134-146.
 Johnson, J. A. (1997). Units of analysis for the description and explanation of personality. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 73-93). San Diego, CA: Academic.
 Kuhlman, D., Zuckerman, M., Gomà-i-Freixanet, M., & Shiomí, K. (2003). *Cultural and gender differences in personality within three cultures*. Manuscript submitted for publication.
 McCrae, R. R. (1982). Consensual validation of personality traits: Evidence from self-reports and ratings. *Journal of Personality and Social Psychology*, 43, 293-303.
 McCrae, R. R., & Costa, P. T., Jr. (1983). Joint factors in self-reports and ratings: Neuroticism, extraversion and openness to experience. *Personality and Individual Differences*, 4, 245-255.
 McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the Five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52, 81-90.
 McCrae, R. R., & Costa, P. T., Jr. (1989). Different points of view: Self-reports and ratings in the assessment of personality. In J. P. Forgas & M. J. Innes (Eds.), *Recent advances in social psychology: An international perspective* (pp. 429-439). Amsterdam: Elsevier.
 Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
 Montag, I. (2001). *The Hebrew ZKPQ: Preliminary reliability and validity results*. Unpublished manuscript.
 Mutén, E. (1991). Self-reports, spouse ratings, and psychophysiological assessment in a behavioural medicine program: An application of the Five-factor model. *Journal of Personality Assessment*, 57, 449-464.
 Norman, W. T., & Golberg, L. R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology*, 4, 681-691.
 Ostendorf, F., & Angleitner, A. (1994). A comparison of different instruments proposed to measure the Big Five. *European Review of Applied Psychology*, 44, 45-53.
 O'Sullivan, D., Zuckerman, M., & Kraft, M. (1996). The personality of prostitutes. *Personality and Individual Differences*, 21, 445-448.

- O'Sullivan, D., Zuckerman, M., & Kraft, M. (1998). Personality characteristics of male and female participants in team sports. *Personality and Individual Differences*, 25, 119–128.
- Prieto, L., Lamarca, R., & Casado, A. (1998). La evaluación de la fiabilidad en las observaciones clínicas: El coeficiente de correlación intraclase [The assessment of reliability in clinical observations: The intraclass correlation coefficient]. *Medicina Clínica*, 110, 142–145.
- Shiomi, K., Kuhlman, D., Zuckerman, M., Joireman, J., Sato, M., & Yata, S. (1996). Examining the validity and reliability of a Japanese version of the Zuckerman–Kuhlman Personality Questionnaire (ZKPQ). *Hyogo University of Teacher Education Journal*, 2, 1–13.
- Watson, D., & Clark, L. A. (1991). Self- versus peer ratings of specific emotional traits: Evidence of convergent and discriminant validity. *Journal of Personality and Social Psychology*, 60, 927–940.
- Wu, Y., Wang, W., Du, W., Li, J., Jiang, X., & Wang, Y. (2000). Development of a Chinese version of the Zuckerman–Kuhlman Personality Questionnaire: Reliabilities and gender/age effects. *Social Behaviour and Personality*, 28, 241–250.
- Zuckerman, M. (1991). *Psychobiology of personality*. New York: Cambridge University Press.
- Zuckerman, M. (1992). What is a basic factor and which factors are basic? Turtles all the way down. *Personality and Individual Differences*, 13, 675–681.
- Zuckerman, M. (2002). Zuckerman–Kuhlman Personality Questionnaire (ZKPQ): An alternative five-factorial model. In B. De Raad & M. Perugini (Eds.), *Big Five assessment* (pp. 377–396). Seattle, WA: Hogrefe & Huber.
- Zuckerman, M., & Kuhlman, D. (1993). *Norms for the Zuckerman–Kuhlman Personality Questionnaire (ZKPQ)*. Unpublished manuscript.
- Zuckerman, M., Kuhlman, D., & Camac, C. (1988). What lies beyond E and N? Factor analyses of scales believed to measure basic dimensions of personality. *Journal of Personality and Social Psychology*, 54, 96–107.
- Zuckerman, M., Kuhlman, D., Joireman, J., Teta, P., & Kraft, M. (1993). A comparison of three structural models for personality: The Big Three, the Big Five, and the Alternative Five. *Journal of Personality and Social Psychology*, 65, 757–768.
- Zuckerman, M., Kuhlman, D., Thornquist, M., & Kiers, H. (1991). Five (or three): Robust questionnaire scale factors of personality without culture. *Personality and Individual Differences*, 12, 929–941.

Montserrat Gomà-i-Freixanet
 Department of Health Psychology
 Autonomous University of Barcelona
 Edifici B
 E-08193 Bellaterra
 Catalonia, Spain
 Email: montserrat.goma@uab.es

Received April 27, 2004
 Revised September 23, 2004