

## The Shortened Person's Relating to Others Questionnaire (PROQ3): Comparison of the Internet-Administered Format With The Standard- Written One Across Four National Samples

Argyroula E. Kalaitzaki  
Technological Educational Institute of Crete

John Birtchnell  
Institute of Psychiatry, London

Sean Hammond  
University College Cork

Cor De Jong  
Radboud University

The study aims to validate the shortened version of the Person's Relating to Others Questionnaire (PROQ3), a measure of negative and maladaptive relating to others, for data collection via the Internet across 4 national samples. The psychometric properties of the Internet-administered (IN) format of the PROQ3 in 4,802 participants (169 British; 360 Irish; 1,110 Dutch; and 3,163 Greek) were compared with that of the standard-written (SW) version in 1,918 participants (338 British, 403 Irish, 204 Dutch, and 973 Greek), in respect of its measurement and structural equivalence. Internal consistency, as estimated by alpha coefficient and item-to-scale homogeneity, were consistently acceptable across nationality and modality. There was agreement in regard to the interscale correlations across nationality and modality. Lower mean scores for the British sample in the SW format, and lower mean scores for the Irish sample in the IN format were found. The structural equivalence across modality and nationality was also supported: A consistent 8-factor underlying structure, as supported by a multiple group factor analysis, and an octagonal higher order, as supported by a 3-way multidimensional scaling procedure, were found. It was concluded that the PROQ3 can be administered via the Internet with maintained psychometric properties for clinicians to screen people with interpersonal relating deficiencies and for researchers to collect data.

*Keywords:* interpersonal relationships, psychometric properties, measurement and structural equivalence, multiple group factor analysis, multidimensional scaling procedure

*Supplemental materials:* <http://dx.doi.org/10.1037/pas0000057.supp>

A greater internationalization of psychological assessment has resulted in an increasing use of Internet-administered assessment, which, in turn, serves to broaden the wide-scale access to data across national borders (Naglieri et al., 2004; The International Test Commission [ITC], 2006). The Internet-administered collection of data by self-report questionnaires is becoming increasingly popular, which is due to its many advantages over standard-written (SW) surveys. These advantages include cost and time benefits, ease of administration, automated scoring, and recruitment of large numbers of participants (Gosling, Vazire, Srivastava, & John,

2004; Reips, 2000). A wide range of studies, extending beyond simple surveys, are now being carried out in this way. Some measures have been developed specifically for Internet use, whereas others, probably the majority, are Internet-administered (IN) versions of existing SW measures.

An assumption underlying the use of the Internet is that this method of data elicitation does not significantly bias responses, so the SW format produces scores compatible with those from Internet sources. Further, the conversion of a previously valid, reliable, and psychometrically sound SW measure to IN format, could be expected to lead to these properties transferring to the IN version. However, some investigators have acknowledged that the results obtained by the IN method may not readily compare with those obtained by more conventional approaches, owing to bias of a lower response rate and differences in respondents' demographic characteristics (Bech & Kristensen, 2009). Other sources of bias include greater Internet familiarity of certain groups of participants, the potential lack of control of conditions in which the questionnaires are completed, and the unknown motives of the participants. All of these potential sources of bias may threaten the quality and validity of data obtained through IN (Reips, 2000). Thus, equivalence between the two

---

This article was published Online First January 5, 2015.

Argyroula E. Kalaitzaki, Social Work Department, Technological Educational Institute of Crete, Crete, Greece; John Birtchnell, Institute of Psychiatry, London, England; Sean Hammond, Department of Applied Psychology, University College Cork, Cork, Ireland; and Cor De Jong, Department of Clinical Psychology, Radboud University, Nijmegen, the Netherlands.

Correspondence concerning this article should be addressed to Argyroula E. Kalaitzaki, Social Work Department, Health and Social Welfare School, Technological Educational Institute of Crete, Stavromenos, PC 71004, Heraklion, Crete, Greece. E-mail: [akalaitzaki@staff.teicrete.gr](mailto:akalaitzaki@staff.teicrete.gr)

formats must be established rather than assumed (Buchanan, 2002, 2003).

To date there have been numerous studies demonstrating relative equivalence between SW and IN questionnaires used in a range of research areas, such as personality (see Buchanan, Paine, Joinson, & Reips, 2007, for review) and psychopathology (Hedman et al., 2010). Although relative equivalence over modalities has already been established for such measures, it has yet to be examined for measures of interpersonal relating. To date, only two studies are known to have attempted to validate some of the most well-known interpersonal measures in this way. Hamby, Sugarman, and Boney-McCoy (2006) have investigated the psychometric equivalence across modality for the Revised Conflict Tactics Scales, a measure of partner violence. Brock, Barry, Lawrence, Dey, and Rolffs (2012) assessed such interpersonal constructs as relationship satisfaction, communication/conflict management, partner support and emotional intimacy together with intrapersonal constructs through self-report questionnaires. The measures were generally shown to be reliable when administered via the Internet and were quantitatively and qualitatively equivalent across modality. Controversial results, even by the same authors (Kraut et al., 1998, 2002), on whether Internet users have an increased rate of depression, maladjustment, or social isolation, necessitates further work on the validation of data collection through IN.

The present study seeks to determine psychometric equivalence of the two sampling methodologies across four national samples. Psychological assessments are being increasingly developed for international use, and this requires that measurement bias due to national or cultural milieu be identified and minimized (de Klerk, 2008; Van de Vijver & Tanzer, 1997). The use of such assessments across nationalities carries an implicit assumption of a degree of universality in the constructs being measured and this has been a guiding principle in the cross-national assessment of personality (Hofstede & McCrae, 2004; Piedmont & Chae, 1997). However, critics have observed that such an assumption may not be appropriate as the very nature of personality may be influenced by the culture from which the individual emerges (Allik & Realo, 2009). Individualistic or collectivistic cultural values, determining an individual's relationship with his or her group, may play a significant role (Triandis, 1994). It may then be reasonable to assume that the concept of *relating to others* is likely to be highly culturally determined given the social context of the domain.

The need in clinical psychology and psychotherapy to assess a person's less-than-competent relating tendencies toward other people (negative relating) led to the development of the 96-item Person's Relating to Others Questionnaire (PROQ; Birtchnell, Falkowski, & Steffert, 1992). A revised version of the PROQ was produced, called the PROQ2 (Birtchnell & Shine, 2000), which intended to improve the clarity and factorial structure and to reduce the correlation between scales. It is now in its third iteration (Birtchnell, Hammond, Horn, De Jong, & Kalaitzaki, 2013), which contains half the items of the previous versions. It is based on relating theory (Birtchnell, 1993/1996), which proposes that relating can be classified within a biaxial model constructed around the poles of two intersecting axes: a horizontal one concerning *closeness* (seeking to be involved) versus *distance* (seeking to be separate) and a vertical one concerning *upperness* (relating from a position of strength) and *lowerness* (relating from a position of weakness). These poles are called the states of relatedness or the

relating objectives. All four of them are considered to be equally advantageous, and ideally, during the course of maturation, people should acquire the ability to attain and hold on to them. People vary in this ability. For any pole, competent relating is called *positive* and less-than-competent relating is called *negative*, and it is negative relating that the PROQ was designed to measure.

The model has been expanded by introducing intermediate positions between the four main ones, and the characteristics of each intermediate position are a blending of those of the positions to either side of it. This creates eight positions which together are called the *interpersonal octagon* (Birtchnell, 1993/1996). Each position has been given a two-word name, the first word referring to the vertical axis, the second referring to the horizontal axis. For the main positions the word *neutral* is placed where the name for the other axis would have been. The eight scales of the PROQ correspond to the negative relating within the eight positions of the octagon. Moving round the octagon, the names are upper neutral (UN), upper close (UC), neutral close (NC), lower close (LC), lower neutral (LN), lower distant (LD), neutral distant (ND), and upper distant (UD). Both the positive and negative relating characteristics of each of the eight octants have been fully described elsewhere (Birtchnell, 1993/1996, 2002a). Summary definitions of each are shown in Figure 1, depicted in two separate octagons.

The psychometric properties of the SW format of the PROQ and the revised PROQ2 have been extensively studied in English clinical and nonclinical samples (Birtchnell, 2002b; Birtchnell & Evans, 2004), in two English forensic samples (Birtchnell, Shuker, Newberry, & Duggan, 2009), and in two Greek nonclinical samples (Kalaitzaki & Nestoros, 2003; Kalaitzaki, Birtchnell, & Kritsotakis, 2010). The psychometric properties of the latest version (the PROQ3) have been replicated in an English forensic sample (Newberry & Birtchnell, 2011) and within four national samples (British, Irish, Dutch, and Greek), though again only in its SW method (Birtchnell et al., 2013). There was reasonably good general agreement across samples and studies. For instance, alpha coefficients and interscale correlations of the PROQ3 were found to be consistently acceptable across samples. For the psychotherapy sample alpha coefficients and mean scores of the majority of the scales were all significantly higher than for the normative samples, as they had been for the PROQ2 (Birtchnell et al., 2013). The psychometric properties of the IN version have yet to be evaluated.

### Aims of the Study

The aim of the present study was to examine whether the IN version of the PROQ3 has comparable psychometric properties with the validated SW version (Birtchnell et al., 2013) across four different national samples. First, we examined comparability between descriptive psychometric parameters such as internal consistency, item-to-scale homogeneity, and mean score differences between modality and nationality. Second, we addressed the issue of construct validity and, in particular, the structural equivalence in respect of its multidimensional and octagonal structure, across both nationality and modality. The expectation was that the PROQ3 would be reasonably valid and reliable and that there would be evidence for both an eight-factor underlying structure and an octagonal structure, across both nationality and modality. Should psychometric equivalence be demonstrated, it would sup-

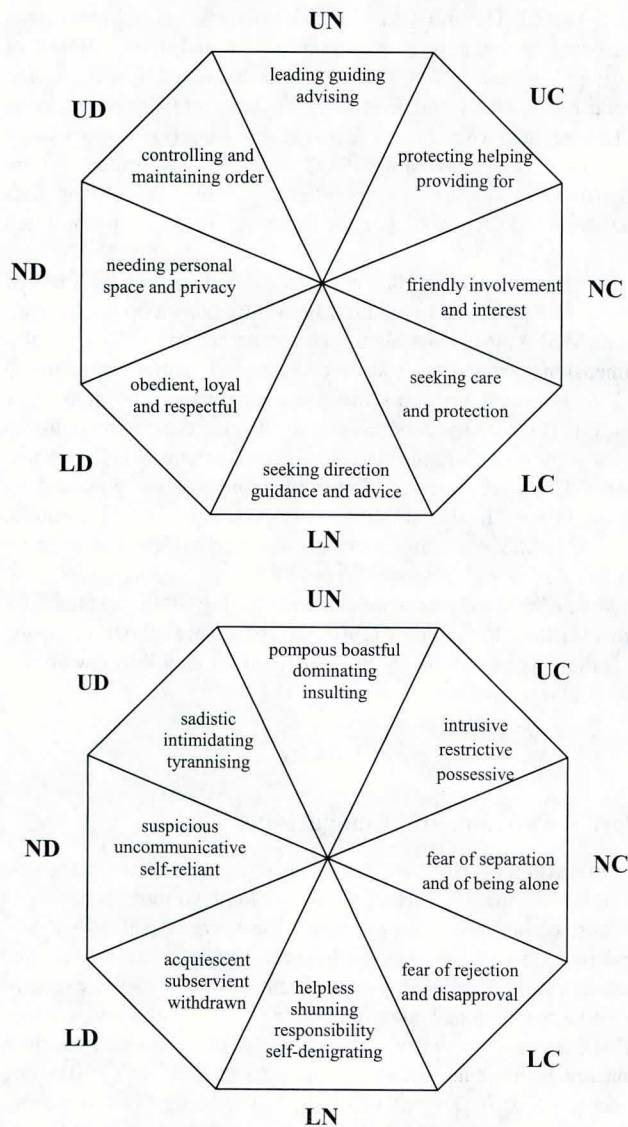


Figure 1. Positive (upper diagram) and negative (lower diagram) forms of relating for each octant, which are labeled as follows: upper neutral (UN), upper close (UC), neutral close (NC), lower close (LC), lower neutral (LN), lower distant (LD), neutral distant (ND), and upper distant (UD). Adapted from Birtchnell, J. (1994).

port the position that both modalities may be used interchangeably in clinical practice.

## Method

### Study Groups/Participants

The study draws on four IN samples: (a) a sample of 3,163 Greek participants (46.2% men and 53.8% women), (b) a sample of 1,110 Dutch participants (30.6% men and 69.4% women), (c) a sample of 169 British participants (60.9% men and 39.1% women), and (d) a sample of 360 Irish participants (17.0% men and 83.0% women). In all sites, the age group of the majority of

the participants was under 25 (39.1%, 49.7%, and 73.9%, respectively), except the Dutch one, where the majority of the participants belonging to the age group 36 to 55 (53.2%). SW data were collected from the same four nations. These samples have been reported on in Birtchnell et al. (2013). They were (a) 973 Greek university students (27.5% men and 72.5% women; response rate 97.7%) of the Technological Education Institute of Crete, with the majority belonging to the age group under 25 (94.1%); (b) 204 participants (41.7% men and 58.3% women; response rate 91.1%) from a community survey carried out in Nijmegen, the Netherlands, with the majority belonging to the age group 36 to 55 (82.4%); (c) 313 British participants from a community survey (50.5% men and 49.5% women, no age has been recorded; response rate nearly 90%); and (d) 403 participants (66.5% men and 33.5% women; response rate nearly 90%) recruited from a community survey carried out in Cork, in the Republic of Ireland, with the age of the majority being either under 25 (35.5%) or over 55 (39.7%).

### The Measure

The shortened version of the PROQ3 (Birtchnell et al., 2013) is the latest revision of the PROQ (Birtchnell, Falkowski, & Steffert, 1992). It comprises 48 items (6 items per octant scale, 5 of which are negative; only the negative items are scored), answered in a 4-point Likert-type scale, where 1 = *nearly always true*, 2 = *quite often true*, 3 = *sometimes true*, and 4 = *rarely true*. Like all previous versions, the PROQ3 is scored by computer, and the scores can be represented both numerically, as a list of octant scores, and graphically, as shaded areas of octants in an octagon. Typical items for the eight PROQ3 scales are as follows: *I try to arrange things so that people do what I want* (UN), *I keep a firm hold on someone who is close to me* (UC), *I have a tendency to cling to people* (NC), *I have a dread of being rejected* (LC), *I prefer it when someone else is in control* (LN), *I easily give in to people* (LD), *I do not let people get too close to me* (ND), *I tend to get back at people who offend me* (UD). Each scale has a maximum score of 15. The maximum total score of the questionnaire is 120. It can be downloaded from <http://www.johnbirtchnell.co.uk>. For both the Dutch and Greek samples, we used translations. These were carried out by bilingual English-speaking psychologists. Back translations were made by an independent English-speaking colleague, who had not seen the original version. Discrepancies were discussed with the English originator of the questionnaire, and the final version was arrived at by mutual agreement. Validation of the Dutch version of the PROQ3 was achieved through its comparison with two interpersonal measures, the Interpersonal Check List-Revised (ICL-R) and the Inventory of Interpersonal Problems-Circumplex Scales (IIP-C) (Horn, 2004). It was found that all scales of all questionnaires could be predicted through their theoretically equivalent scales of one of the other measures. It was concluded that all the three instruments seem to measure the same dimensions of interpersonal behavior and there are more resemblances than differences in between the three.

### Study Procedure/Data Collection

For the elicitation of the IN data, the website was set up initially in English and subsequently it was translated into Greek and

Dutch. The samples were recruited using the snowball or chain-sampling technique. Initially, an electronic message was sent from the investigators to all their contacts. In addition, the Greek investigator sent an e-mail to each person (i.e., teaching, administrative and technical staff, students) registered in the e-mail directory of Greek universities and technological education institutes that were publicly available. The message contained the description of the study, issues of anonymity and confidentiality, and information about the availability of the scores (i.e., a person's negative relating style), both graphically by way of an octagonal diagram and numerically in the form of a frequency distribution, upon completion of the questionnaire as a reward for participation. This was typically the informed consent page, which preceded the questionnaire. Finally, it invited the readers to participate by clicking on the hypertext link that was embedded in the text and pass the message on to their own contacts. Linked to the web page, an introductory page briefly described the PROQ3, and then the participant was presented with the demographics (i.e., gender and age) and the questionnaire. The questionnaire's resemblance to the SW (divided into two pages) was intended to control for the confounding effects of differences in administrative procedure. No questionnaires with missing data could be submitted. (This "missing data checking" is another benefit of computerized testing, including web surveys.) The Irish Internet survey was independently constructed using SurveyMonkey (Palo Alto, California, USA). The instructions and information provided were similar to that set up for the other samples, with the exception that no access to the final scores was available.

The SW sample were recruited by a variety of means. For the Greek, Dutch, and Irish samples, ethical approval was granted by the School ethics committee. Respondents were required to be over 18 years. The Greek questionnaires were filled out anonymously by university students during regularly scheduled classes. The British, Irish, and Dutch surveys were distributed using a snowball sampling technique. For the Irish sample, the first tier of respondents were university students, whereas for the British and Dutch samples, the first tier of respondents were mostly acquaintances of the investigators; for all samples the subsequent levels consisted of acquaintances and relatives of the first tier. All responses were returned to the principal investigator anonymously in envelopes provided with each questionnaire. Part of the Dutch questionnaires was also distributed to the neighborhood. Preceding the questionnaire, participants from all sites were informed that the study was voluntary, that their results would be kept confidential, and that they could withdraw from the study at any time. There was no financial reward for their participation.

### Data Analysis

Initially correlational analysis was carried out to explore the consistency of the interscale structures across nationality and modality. Differences between the mean correlation coefficients across modality were examined by applying Fisher's *r*-to-*z* transformation. Assessment of the internal consistency of the scales was carried out and comparisons between the modality were facilitated by using Feldt's (1969) test for independent alpha coefficients. The mean inter-item coefficients for the eight PROQ3 subscales for both the IN and SW formats for all sites were also calculated.

A  $4 \times 2$  factorial multivariate analysis of variance was performed to examine the independent and joint effects of *nation* (1 = Greek, 2 = Dutch, 3 = British, and 4 = Irish) and *modality* (1 = IN and 2 = SW) on the eight PROQ3 scores. Likewise, a  $4$  (Nation)  $\times$   $2$  (Modality) analysis of variance (ANOVA) on the overall PROQ3 score was performed. Where significant multivariate effects were revealed, follow-up univariate ANOVAs were conducted using Bonferroni-corrected alpha levels ( $.05/8 = .006$ ).

For the evaluation of the structural equivalence of the PROQ3 across nationality and modality, both multiple group factor analyses (MGFA) to test its eight-factor structure and a 3-way multi-dimensional scaling analyses to evaluate its octagonal structure in a 2-dimensional common space, were carried out. Common space analysis (COSPA) was developed to identify the commonality in the structure of multiple data sets (Schonemann, 1972; Schonemann, James, & Carter, 1978) built upon a model proposed by Horan (1969). In this way, the ordinal octagon structure can be verified without needing to make assumptions for the stronger circumplex model.

Most of the analyses were performed using SPSS (Version 20), however the MGFA and COSPA analyses were carried out using programs written by one of the authors to run in a Windows-based environment (available upon request).

## Results

### Part 1. Psychometric Comparisons

**Interscale correlations.** The examination of the interscale correlations, the structure of which is examined more precisely in a later section, was used as a test of convergent validity. It was apparent that there was a between-nation (nationality) and between-formats (modality) agreement in regard to the expected positive correlations between adjoining octant scales and the negative correlation between certain scales. The positive correlations concerned the polar scales and the intermediate scales adjoining them (e.g., UN/UD, UN/UC, LN/LD). The highest positive correlations were between the three close scales (UC/NC, UC/LC, and NC/LC). In general, the negative correlations were between UN/LN, LC/UD, LN/UD, and LD/UD. (The detailed results for each nation are available upon request.) No difference between the mean correlation coefficients by applying Fisher's *r*-to-*z* transformation was found between the two formats for any site (Greek:  $z = -1.44, p = .150$ ; Dutch:  $z = -0.14, p = .889$ ; British:  $z = -.76, p = .447$ ; Irish:  $z = 1.15, p = .250$ ).

**Reliability.** The alpha coefficients for each scale across nations and administration formats are presented in Table 1. It can be seen that, where statistically significant differences are found, the IN format manifests a greater degree of internal consistency than does the SW format. Nevertheless, as estimates of lower bound reliability, the alpha coefficients observed in this data reveal reasonable levels of internal consistency with few exceptions falling below 0.70. In addition, it was observed that the mean inter-item coefficients for the eight PROQ3 subscales for both the IN and SW formats for all sites were within the acceptable range of .10 and .50 for multifactor scales (Briggs & Cheek, 1986). These added to the item-to-scale homogeneity.

Table 1  
Psychometric Properties of the PROQ3 Scores by Nationality and Modality

	Greek			Dutch			British			Irish		
	IN	SW	w	IN	SW	w	IN	SW	w	IN	SW	w
UN												
M	8.33	6.70		7.00	7.23		7.43	7.76		7.06	7.82	
SD	3.21	3.35		2.76	3.85		3.60	3.94		3.32	3.89	
$\alpha$	.75	.74	.94	.72	.70	.92	.75	.70	.84	.76	.70	.79
UC												
M	4.23	5.04		2.32	3.88		3.60	4.90		9.53	4.66	
SD	3.29	3.08		2.36	3.27		3.18	4.15		3.75	3.89	
$\alpha$	.77	.64	.64 <sup>a</sup>	.75	.71	.89	.80	.80	.99	.85	.77	.64 <sup>a</sup>
NC												
M	6.05	7.27		3.16	5.26		4.01	5.50		8.75	4.63	
SD	3.26	3.33		2.22	4.23		3.06	4.04		3.73	3.52	
$\alpha$	.73	.70	.89	.64	.60	.90	.75	.75	.97	.80	.68	.62 <sup>a</sup>
LC												
M	6.96	6.85		4.58	5.59		6.22	9.42		5.94	5.70	
SD	3.55	3.39		3.19	3.88		3.72	4.71		3.79	3.94	
$\alpha$	.76	.68	.75 <sup>a</sup>	.80	.75	.80	.81	.84	.88	.82	.72	.66 <sup>a</sup>
LN												
M	5.41	5.34		4.30	5.63		4.89	5.86		8.92	5.55	
SD	2.97	3.00		2.58	4.12		2.73	4.21		3.31	3.82	
$\alpha$	.74	.67	.80 <sup>a</sup>	.78	.80	.92	.78	.81	.84	.80	.71	.68 <sup>a</sup>
LD												
M	5.62	5.27		5.45	4.72		5.53	7.81		7.35	5.57	
SD	2.67	2.84		2.93	3.97		2.99	4.23		3.30	3.39	
$\alpha$	.53	.56	.94	.72	.69	.92	.66	.73	.80	.67	.59	.82
ND												
M	6.74	7.43		4.77	3.72		6.18	7.64		6.93	6.72	
SD	3.39	3.19		2.82	2.90		3.61	4.60		3.57	4.16	
$\alpha$	.78	.71	.75 <sup>a</sup>	.73	.78	.80	.84	.81	.85	.81	.75	.74 <sup>a</sup>
UD												
M	6.41	6.28		3.67	2.89		6.34	6.02		7.58	6.53	
SD	3.17	3.04		2.62	3.25		3.04	4.10		3.10	3.63	
$\alpha$	.67	.60	.82 <sup>a</sup>	.69	.59	.76 <sup>a</sup>	.69	.74	.82	.70	.64	.83
Total												
M	49.76	49.95		35.25	38.94		44.19	54.90		61.89	48.45	
SD	14.37	15.36		11.45	15.83		14.16	16.80		16.78	14.25	
$\alpha$	.86	.86	.99	.84	.83	.98	.86	.82	.77	.88	.82	.65 <sup>a</sup>

Note. The w statistic is Feldt's test for testing the equality of independent alpha coefficients. PROQ3 = shortened version of the Person's Relating to Others Questionnaire; IN = Internet-administered format; SW = standard-written format; UN = upper neutral; UC = upper close; NC = neutral close; LC = lower close; LN = lower neutral; LD = lower distant; ND = neutral distant; UD = upper distant.

<sup>a</sup> Statistical significance at the .006 level, following Bonferroni adjustments.

**Score differences as a function of the nation and format.** A significant Nation  $\times$  Modality interaction effect on the eight PROQ3 scales was found (Wilks'  $\lambda = 0.820$ ;  $F[24, 19055.6] = 56.180$ ,  $p < .0001$ ; partial  $\eta^2 = 0.064$ ). Univariate effects were significant for all eight scales. For the Irish and Greek samples, mean scores on certain scales of the IN format were significantly higher than that of the SW format, whereas the opposite was true for the British and Dutch samples. The two-way ANOVA also revealed a significant Nation  $\times$  Modality interaction effect,  $F(3) = 88.426$ ,  $p < .0001$ , partial  $\eta^2 = .039$ , for the total scores. For the Irish sample, UC, NC, LN, UD, and total scale scores for the IN method were higher than that of the SW method; whereas for the Greek sample, only the UN scale score was higher. For the British sample, LC, LD, ND, and total scale scores for SW method were higher than that of the IN method; whereas for the Dutch sample, only the LN score was higher. These effects are presented graphically in Figure 2.

## Part 2. Structural Equivalence

**Eight-factor structure.** The statistical analyses reported here demonstrate that PROQ3 scores do vary across modality of presentation and nationality, although not in any clearly consistent manner. This does not permit judgments of whether the essential meaning of the measures is changed by modality, which may then result in statistical comparisons with little psychological sense. To address this issue, an evaluation of the structural equivalence of the PROQ3 across both nationality and modality was carried out using a multiple group factor analytic procedure. The underlying model of the PROQ3 is considered to be of eight scales arranged in an octagon. This basic model may be expressed as an oblique eight-factor structure with order constraints on the factor covariances indicative of an ordinal octagon structure. We tested the model by expressing the hypothesized eight-factor structure as a zero-unity target matrix and carried out an oblique multiple group factor

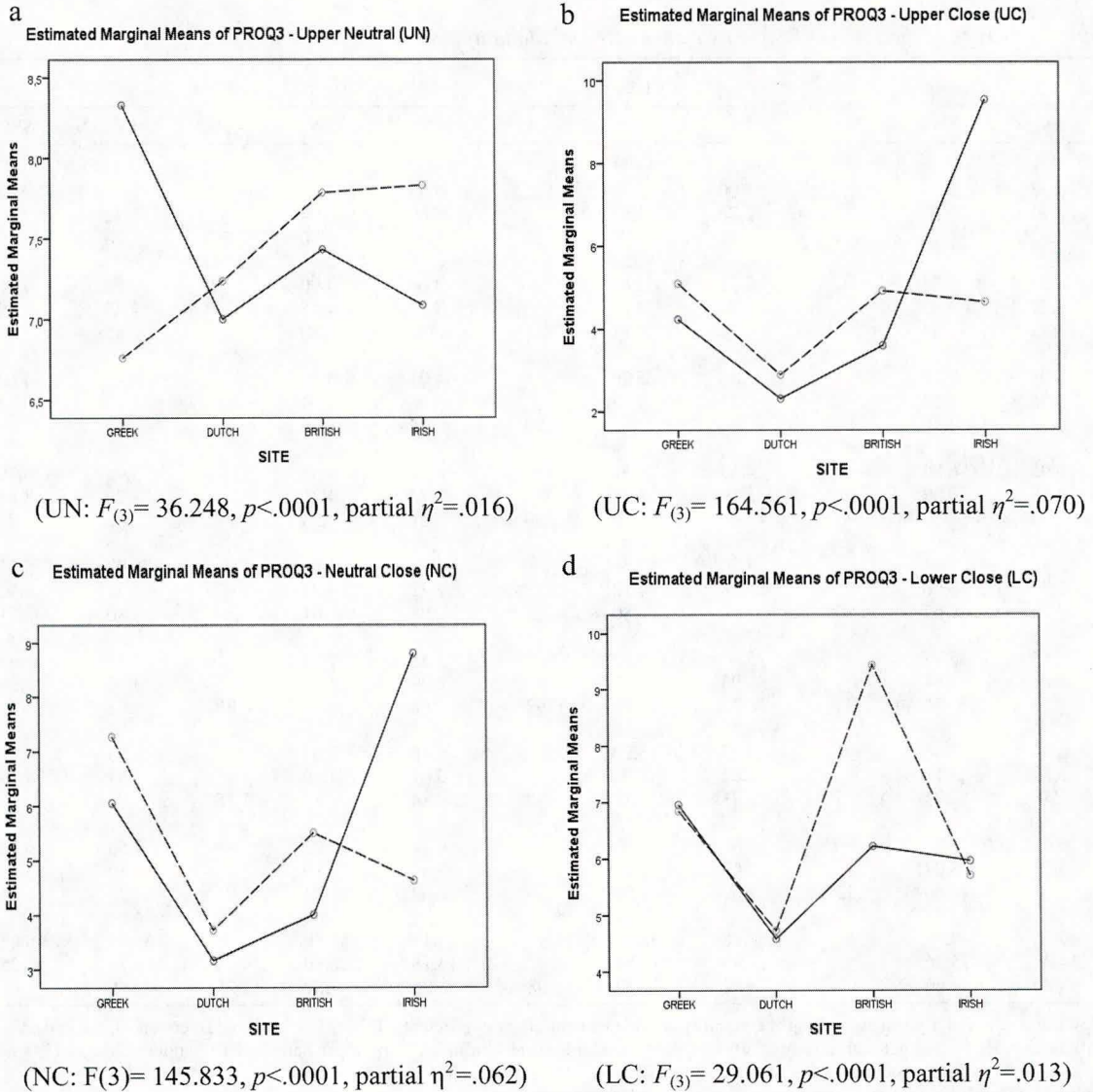
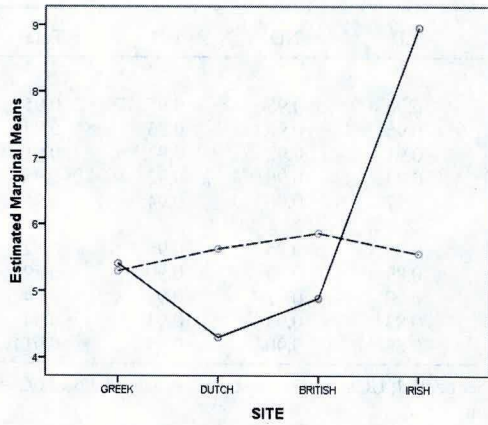


Figure 2. Mean scores for the eight PROQ scales broken down by nationality and method. Continuous line represents Internet-administered method (IN), whereas dot line represents standard-written method (SW). Octants are labeled as follows: upper neutral (UN), upper close (UC), neutral close (NC), lower close (LC), lower neutral (LN), lower distant (LD), neutral distant (ND), and upper distant (UD).

analysis (Levin, 1988) for each of the eight data sets broken down by nationality (4) and modality (2). To assess the fit of the model to each data set, we used Fleming's (1985) index. This is a simple signal-to-noise ratio derived from Kaiser's (1974) simplicity index and is independent of any assumptions of the latent distribution. It also has the advantage of allowing for a fit for each factor and each variable and provides an overall estimate of fit. A limitation of all such confirmatory analyses is that a good fit does not imply best fit. To this end, the current analysis uses a Monte Carlo method of generating 10,000 randomly generated models. We then compared the hypothesized model to the distribution of randomly generated models to provide an approximate estimate of model confirmation (Curtin & Hammond, 2012; Hammond, O'Rourke, Kelly, Bennett, & O'Flynn, 2012). (The full factor pattern results are available upon request.) The results are summarized in Table 2.

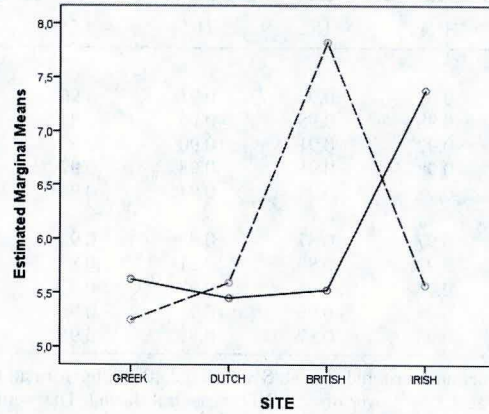
The overall fit for each data set to the hypothesized eight-factor model ranged from 0.89 to 0.95. When transformed to  $z$  values, using the randomly generated norms, it was clear that the total fit indices are all over five standard deviations above the random expected value, as demonstrated in the last column of Table 2. In practice, the Fleming indices reveal a reasonably symmetrical distribution over the 10,000 random models, but there is no statistical justification for assuming they will manifest a normal distribution. Therefore, the interpretation of  $z$  should be considered instructive rather than hypothesis testing. Nevertheless, it is quite apparent that, for every analysis, Birtchnell's (1993/1996) eight-factor model fit the data extremely well and there are no grounds to suppose that there is any appreciable bias due to nationality or administration method. The fit indices for each separate factor also revealed excellent fit ranging from 0.83 to

e Estimated Marginal Means of PROQ3 - Lower Neutral (LN)



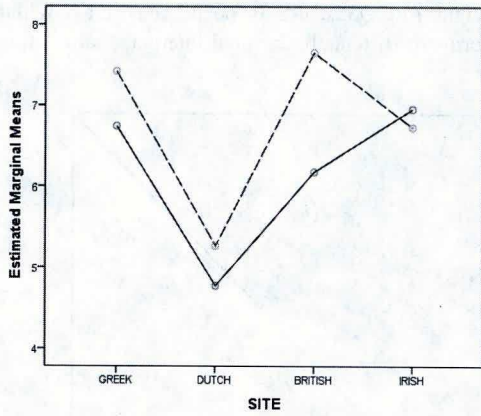
(LN:  $F_{(3)} = 85.295, p < .0001, \text{partial } \eta^2 = .037$ )

f Estimated Marginal Means of PROQ3 - Lower Distant (LD)



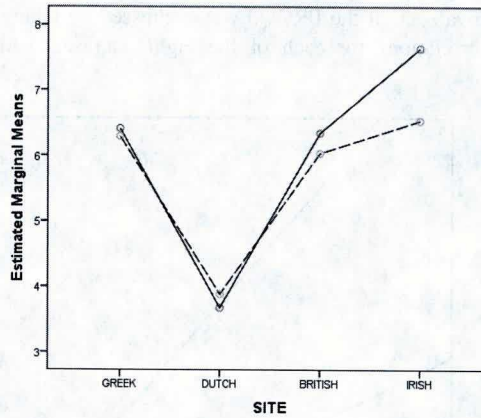
(LD:  $F_{(3)} = 46.304, p < .0001, \text{partial } \eta^2 = .021$ )

g Estimated Marginal Means of PROQ3 - Neutral Distant (ND)



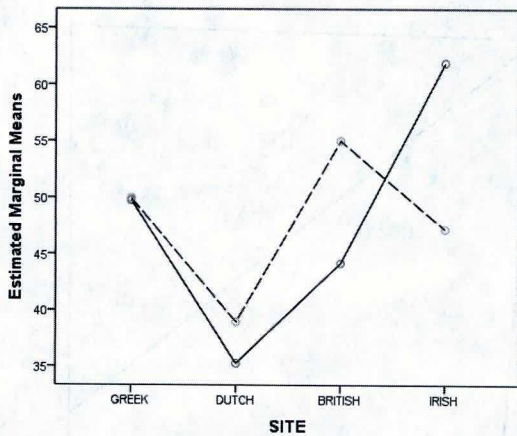
(ND:  $F_{(3)} = 6.166, p < .0001, \text{partial } \eta^2 = .003$ )

h Estimated Marginal Means of PROQ3 - Upper Distant (UD)



(UD:  $F_{(3)} = 6.165, p < .0001, \text{partial } \eta^2 = .003$ )

i Estimated Marginal Means of PROQ3 - Total score



(Total:  $F_{(3)} = 88.426, p < .0001, \text{partial } \eta^2 = .039$ )

Figure 2. (continued)

Table 2  
Fit Indices of the Eight Data Sets to the Common Hypothesis Matrix

	UN	UC	NC	LC	LN	LD	ND	UD	Total	z
IN										
Greek	0.97	0.94	0.91	0.96	0.92	0.96	0.95	0.95	0.95	13.38
Dutch	0.98	0.98	0.96	0.96	0.91	0.96	0.96	0.95	0.95	9.89
British	0.92	0.91	0.90	0.90	0.87	0.91	0.92	0.92	0.91	12.23
Irish	0.98	0.95	0.98	0.97	0.92	0.97	0.96	0.92	0.95	13.38
Combined	0.98	0.94	0.94	0.98	0.93	0.97	0.96	0.94	0.95	9.56
SW										
Greek	0.97	0.87	0.83	0.93	0.93	0.95	0.95	0.95	0.92	8.76
Dutch	0.92	0.89	0.91	0.89	0.84	0.89	0.89	0.90	0.89	8.63
British	0.97	0.94	0.92	0.97	0.95	0.90	0.91	0.97	0.94	10.88
Irish	0.95	0.96	0.95	0.95	0.95	0.93	0.93	0.93	0.94	12.27
Combined	0.95	0.87	0.89	0.95	0.88	0.88	0.90	0.95	0.91	5.49

Note. IN = Internet-administered format; SW = standard-written format; UN = upper neutral; UC = upper close; NC = neutral close; LC = lower close; LN = lower neutral; LD = lower distant; ND = neutral distant; UD = upper distant.

0.98. Therefore, the structural equivalence across method and nationality was supported.

**Higher order octagonal structure.** The question of the higher order equivalence of the PROQ3 was addressed by taking the interscale correlations for each of the eight data sets and

subjecting them to a three-way Multidimensional Scaling Procedure (Borg & Groenen, 2005; Schonemann, 1972; Horan, 1969). The resulting common space plot is presented in Figure 3. Each of the eight PROQ3 scales is positioned in a circular order that conforms to Birtchnell's ordinal interpersonal relating theory and

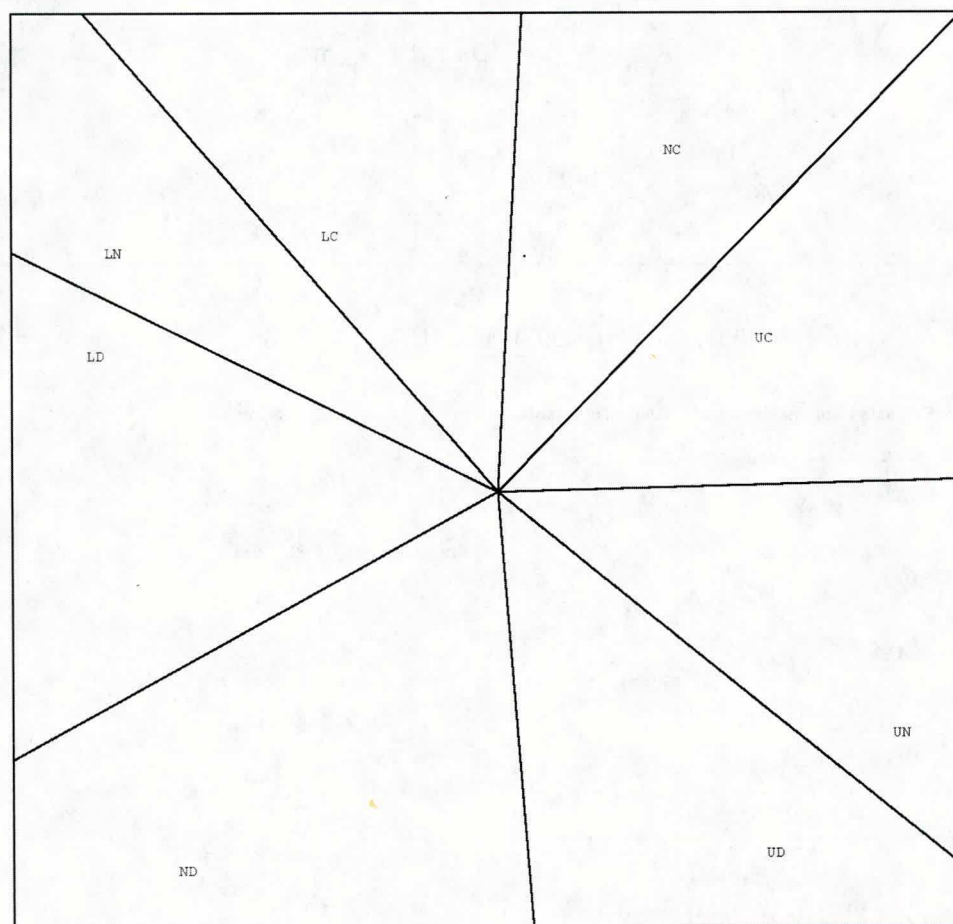


Figure 3. Plot of the 2-Dimensional Common Space for the 8 PROQ3 subscales demonstrating the octagon order.



the space can be sectioned into eight regions specifying his octagon model.

The extent to which this space represents the eight separate data sets was evaluated by examining the common space index and the diagonality index for which perfect fit would produce values of 1.0 and 0.0, respectively. These are presented in Table 3. It is clear that both indices suggest a good fit to this octagon structure, although the British SW set and the Irish IN set show slightly elevated diagonality, which suggests that, to fully maximize fit, some form of differential weighting may be required. Nevertheless, it is clear that the order of the scales within the 2-dimensional space coincide very well with the Octagon structure.

### Discussion

This is the first study to investigate whether the SW and IN versions of the PROQ3, a questionnaire measuring the concept of *relating to others*, yield the same results, particularly across four nations. To the authors' knowledge, no other study has attempted this enterprise for a measure of relating for both nationality and modality. The results were very promising. It has been acknowledged that the PROQ3 could carry the risk of lower reliability because it contains half the number of items per scale than the PROQ2 and somewhat higher measurement errors (Stanton, 1998). Reliability and validity cannot be assumed to remain similar when an SW questionnaire is converted to an IN format, as multiple testing features are changed (Reips, 2000). However this was not the case in the current study. Of note is that, particularly for the Greek and Irish samples, alpha reliabilities appeared in general higher for the IN than for the SW version. It was also reassuring that the coefficients compared reasonably well with those in equivalent samples of the longer PROQ2 (Birtchnell & Evans, 2004). One weakness of the SW version was the low alpha for the revised LD scale (Birtchnell et al., 2013). It was disappointing to find that it remained relatively low in the British, Irish, and Greek samples across both modalities. It has been previously suggested that younger participants have greater difficulty in conceptualizing the LD scale than do older ones (Birtchnell et al., 2013); and these three samples were somehow younger (<25) than was the Dutch (range: 36–55).

Generally, the results were consistent with expectations (Buchanan & Smith, 1999), with the majority of the mean score comparisons between the two formats found to be equivalent across nations. The only exception was that the mean scores on the SW method were significantly higher than those on the IN method

for the British sample; the opposite was true for the Irish sample. A significantly higher rate of disclosing personal information, such as negative affect, in IN surveys than in traditional ones has been reported (Buchanan, 2002; Denissen, Neumann, & van Zalk, 2010; Valkenburg & Peter, 2009), which may justify these national variations. These differences may also be attributed to differences between samples tested either by IN measures or SW ones (e.g., Buchanan, 2003), as well as to cultural differences. Grieve and De Groot (2011) found no significant influence of test administration modality (for either IN or SW measures) in faked responses. Therefore, it could be considered that the means in the present study represent the real scores for these nations. Kalaitzaki and Nestoros (2003) concluded that although the Greek translation of the longer PROQ2 sufficiently compared with the English version, minor differences between the two cultures were observed regarding their relating tendencies. It has been suggested that English-speaking nations, such as England, are mainly individualistic, whereas Greece is a mixture of individualist and profoundly collectivistic values (Green, Deschamps, & Paez, 2005; Hofstede, 1980). Moreover, Schwartz's (2004) cultural value model has suggested that Western European countries are characterized by intellectual autonomy, egalitarianism and harmony, whereas English-speaking countries are characterized by mastery and affective autonomy. One might assume that all these cultural values are closely interwoven with Birtchnell's (1993/1996) concept of an individual's relating to others.

As expected (Birtchnell et al., 2013), high positive correlations between neighboring scales were observed across modality and nationality. This is in line with the understanding that it is possible for one to relate negatively from any number of states of relatedness (octants), even from those on the opposite sides of the octagon (Birtchnell, in press). There were a few negative correlations between the opposite scales, but they were not significant. These results confirm the absence of bipolarity, which is a basic assumption of the octagonal theory. Despite the use of a noncommon language between the samples, which seems plausible to have contributed to some decrease of measurement equivalence (De Beuckelaer, Lievens, & Swinnen, 2007), measurement equivalence was demonstrated. In general, the PROQ3 was shown to be reasonably reliable and valid across nationality and modality.

Methods for evaluating the degree of bias inherent in national or methodological variation are many and varied (Van de Vijver & Leung, 1997), but they mostly come down to an appraisal of structural consistency. In the case of the PROQ3, an a priori eight-factor structure is posited, and furthermore the ordinal arrangement of these factors at a higher order is theoretically determined to conform to an octagon structure. Because the PROQ3 is based on Birtchnell's (1993/1996) relating theory, demonstrating a consistent eight-factor structure, would provide support of this underlying theory. The consistency of this latent structure across both nationality and modality was supported. There was clear evidence for both a consistent eight-factor underlying structure and also an octagonal higher order. This spatial arrangement was clearer and more distinct and unambiguous to that previously reported on the SW format (Birtchnell et al., 2013). In summary these results support the notion of universality (albeit within a European context) in the relating to others structure.

A methodological limitation of the current study is the comparison of different surveys. If participants had been randomly as-

Table 3  
*Fit Indices for the Common Space Analysis*

Data set	Common space index	Diagonality index
Greek SW	0.894	0.053
Greek IN	0.937	0.041
Dutch SW	0.871	0.014
Dutch IN	0.901	0.016
British SW	0.922	0.114
British IN	0.939	0.073
Irish SW	0.887	0.050
Irish IN	0.915	0.128

Note. IN = Internet-administered format; SW = standard-written format.

signed to either the SW or the IN condition first, confounding factors such as sample composition and sampling method may have been avoided. Another evident limitation of the current study is its Eurocentrism. It remains an open question as to whether such findings might apply in countries where the culture is markedly different; though the data presented here are promising and may serve as the basis for further study on the universality of interpersonal relating. It is important not to overgeneralize the findings from the SW versus IN comparisons particularly in certain populations (e.g., clinical) that may behave differently (Fortson, Scotti, Del Ben, & Chen, 2006). Future work on IN measures of relating to others will need to deal with issues such as fraudulent responding and sampling bias (e.g., IN surveys may have favored participants who were young, well-educated, of middle to high socioeconomic status, and had access to Internet).

The implications of these findings are mainly twofold. First, the universality of Birtchnell's (1993/1996) theory of relating is supported, at least in a Western European context. Birtchnell's biaxial system seems to be adequately consistent internationally. Second, the degree of psychometric equivalence manifested supports a recommendation that the PROQ3 may be applied in clinical practice across nationalities using either modality. It could be assumed that the scores obtained from one particular nationality using one modality may be reasonably comparable to scores obtained from another nationality or modality. The increasing use of the Internet for the collection of psychometric data significantly increases the opportunity for obtaining input from international sources. The onus is on test developers to demonstrate that the measures involved are valid and equivalent across such diverse populations. The results of this research project reveal that the PROQ3 demonstrates equivalent psychometric properties across both nationality and modality of data elicitation.

## References

- Allik, J., & Realo, A. (2009). Personality and culture [Editorial]. *European Journal of Personality, 23*, 149–152. <http://dx.doi.org/10.1002/per.718>
- Bech, M., & Kristensen, M. (2009). Differential response rates in postal and web-based surveys among older respondents. *Survey Research Methods, 3*, 1–6.
- Birtchnell, J. (1993/1996). *How humans relate: A new interpersonal theory*. Hove, UK: Psychology Press.
- Birtchnell, J. (1994). The interpersonal octagon: An alternative to the interpersonal circle. *Human Relations, 47*, 518–524. <http://dx.doi.org/10.1177/001872679404700503>
- Birtchnell, J. (2002a). *Relating in psychotherapy: The application of a new theory*. Hove, UK: Brunner-Routledge.
- Birtchnell, J. (2002b). Psychotherapy and the interpersonal octagon. *Psychology and psychotherapy: Theory, research and practice, 75*, 349–363. <http://dx.doi.org/10.1348/147608302320365235>
- Birtchnell, J. (2014). The interpersonal circle and the interpersonal octagon: A confluence of ideas. *Clinical Psychology & Psychotherapy, 21*, 61–72. <http://dx.doi.org/10.1002/cpp.1819>
- Birtchnell, J., & Evans, C. (2004). The Person's Relating to Others Questionnaire (PROQ2). *Personality and Individual Differences, 36*, 125–140. [http://dx.doi.org/10.1016/S0191-8869\(03\)00073-4](http://dx.doi.org/10.1016/S0191-8869(03)00073-4)
- Birtchnell, J., Falkowski, J., & Steffert, B. (1992). The negative relating of depressed patients: A new approach. *Journal of Affective Disorders, 24*, 165–176. [http://dx.doi.org/10.1016/0165-0327\(92\)90064-D](http://dx.doi.org/10.1016/0165-0327(92)90064-D)
- Birtchnell, J., Hammond, S., Horn, E., De Jong, C., & Kalaitzaki, A. (2013). A cross-national comparison of a shorter version of the person's relating to others questionnaire. *Clinical Psychology & Psychotherapy, 20*, 36–48. <http://dx.doi.org/10.1002/cpp.789>
- Birtchnell, J., & Shine, J. (2000). Personality disorders and the interpersonal octagon. *The British Journal of Medical Psychology, 73*, 433–448. <http://dx.doi.org/10.1348/000711200160606>
- Birtchnell, J., Shuker, R., Newberry, M., & Duggan, C. (2009). An assessment of change in negative relating in two male forensic therapy samples using the Person's Relating to Others Questionnaire (PROQ). *Journal of Forensic Psychiatry & Psychology, 20*, 387–407. <http://dx.doi.org/10.1080/14789940802542840>
- Borg, I., & Groenen, P. J. F. (2005). *Modern Multidimensional Scaling: Theory and Applications* (2nd ed.). New York: Springer.
- Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality, 54*, 106–148. <http://dx.doi.org/10.1111/j.1467-6494.1986.tb00391.x>
- Brock, R. L., Barry, R. A., Lawrence, E., Dey, J., & Rolffs, J. (2012). Internet administration of paper-and-pencil questionnaires used in couple research: Assessing psychometric equivalence. *Assessment, 19*, 226–242. <http://dx.doi.org/10.1177/1073191110382850>
- Buchanan, T. (2002). Online assessment: Desirable or dangerous? *Professional Psychology, Research and Practice, 33*, 148–154. <http://dx.doi.org/10.1037/0735-7028.33.2.148>
- Buchanan, T. (2003). Internet-based questionnaire assessment: Appropriate use in clinical contexts. *Cognitive Behaviour Therapy, 32*, 100–109. <http://dx.doi.org/10.1080/16506070310000957>
- Buchanan, T., Paine, C., Joinson, A. N., & Reips, U.-D. (2007). Development of measures of online privacy concern and protection for use on the internet. *Journal of the American Society for Information Science and Technology, 58*, 154–165. <http://dx.doi.org/10.1002/asi.20459>
- Buchanan, T., & Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology, 90*(Pt. 1), 125–144. <http://dx.doi.org/10.1348/000712699161189>
- Curtin, S., & Hammond, S. (2012). Placing dental clinicians within the normative base regarding vicarious response. *European Journal of Dental Education, 16*, 6–11. <http://dx.doi.org/10.1111/j.1600-0579.2011.00700.x>
- De Beuckelaer, A., Lievens, F., & Swinnen, G. (2007). Measurement equivalence in the conduct of a global organizational survey across countries in six cultural regions. *Journal of Occupational and Organizational Psychology, 80*, 575–600. <http://dx.doi.org/10.1348/096317907X173421>
- de Klerk, G. (2008). Cross-cultural testing. In M. Born, C. D. Foxcroft & R. Butter (Eds.), *Online readings in testing and assessment, International Test Commission*. <http://www.intestcom.org/Publications/ORTA.php>
- Denissen, J. J. A., Neumann, L., & Van Zalk, M. (2010). How the Internet is changing the implementation of traditional research methods, people's daily lives, and the way in which developmental scientists conduct research. *International Journal of Behavioral Development, 34*, 564–575. <http://dx.doi.org/10.1177/0165025410383746>
- Feldt, L. S. (1969). A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient 20 is the same for two tests. *Psychometrika, 34*, 363–373. <http://dx.doi.org/10.1007/BF02289364>
- Fleming, J. S. (1985). An index of fit for factor scales. *Educational and Psychological Measurement, 45*, 725–728. <http://dx.doi.org/10.1177/0013164485454002>
- Fortson, B. L., Scotti, J. R., Del Ben, K. S., & Chen, Y.-C. (2006). Reliability and validity of an Internet traumatic stress survey with a college student sample. *Journal of Traumatic Stress, 19*, 709–720. <http://dx.doi.org/10.1002/jts.20165>
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions

- about internet questionnaires. *American Psychologist*, 59, 93–104. <http://dx.doi.org/10.1037/0003-066X.59.2.93>
- Green, E. G. T., Deschamps, J.-C., & Paez, D. (2005). Variation of individualism and collectivism within and between 20 countries: A typological analysis. *Journal of Cross-Cultural Psychology*, 36, 321–339. <http://dx.doi.org/10.1177/0022022104273654>
- Grieve, R., & De Groot, H. T. (2011). Does online psychological test administration facilitate faking? *Computers in Human Behavior*, 27, 2386–2391. <http://dx.doi.org/10.1016/j.chb.2011.08.001>
- Hamby, S., Sugarman, B. B., & Boney-McCoy, S. (2006). Does questionnaire format impact reported partner violence rates? An experimental study. *Violence & Victims*, 21, 507–518. <http://dx.doi.org/10.1891/0886-6708.21.4.507>
- Hammond, S. M., O'Rourke, M., Kelly, M., Bennett, D., & O'Flynn, S. (2012). A psychometric appraisal of the DREEM. *BMC Medical Education*, 12, 2. <http://dx.doi.org/10.1186/1472-6920-12-2>
- Hedman, E., Ljótsson, B., Rück, C., Furmark, T., Carlbring, P., Lindefors, N., & Andersson, G. (2010). Internet administration of self-report measures commonly used in research on social anxiety disorder: A psychometric evaluation. *Computers in Human Behavior*, 26, 736–740. <http://dx.doi.org/10.1016/j.chb.2010.01.010>
- Hofstede, G. (1980). *Culture's consequences: International differences in work related values*. Beverly Hills, CA: Sage.
- Hofstede, G., & McCrae, R. R. (2004). Personality and culture revisited: Linking traits and dimensions of culture. *Cross-Cultural Research: The Journal of Comparative Social Science*, 38, 52–88. <http://dx.doi.org/10.1177/1069397103259443>
- Horan, C. B. (1969). Multidimensional scaling: Combining observations when individuals have different perceptual structures. *Psychometrika*, 34, 139–165. <http://dx.doi.org/10.1007/BF02289341>
- Horn, E. (2004). *Measuring interpersonal behaviour: A comparison of three models* (Unpublished master's thesis). Department of Clinical Psychology, Radboud University, Nijmegen, the Netherlands.
- Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, 15, 336–352. [http://dx.doi.org/10.1016/S0005-7894\(84\)80002-7](http://dx.doi.org/10.1016/S0005-7894(84)80002-7)
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39, 32–36. <http://dx.doi.org/10.1007/BF02291575>
- Kalaitzaki, A. E., Birtchnell, J., & Kritsotakis, E. (2010). The associations between negative relating and aggression in the dating relationships of students from Greece. *Partner Abuse: New Directions in Research, Intervention, and Policy*, 1, 420–442. <http://dx.doi.org/10.1891/1946-6560.1.4.420>
- Kalaitzaki, A. E., & Nestoros, J. N. (2003). The Greek version of the Revised Person's Relating to Others Questionnaire (PROQ2-GR): Psychometric properties and factor structure. *Psychology and Psychotherapy: Theory, Research and Practice*, 76, 301–314. <http://dx.doi.org/10.1348/14760830322362514>
- Kraut, R., Kiesler, S., Boneva, B., Cummings, J., Helgeson, V., & Crawford, A. (2002). Internet paradox revisited. *Journal of Social Issues*, 58, 49–74. <http://dx.doi.org/10.1111/1540-4560.00248>
- Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukopadhyay, T., & Scherlis, W. (1998). Internet paradox. A social technology that reduces social involvement and psychological well-being? *American Psychologist*, 53, 1017–1031. <http://dx.doi.org/10.1037/0003-066X.53.9.1017>
- Levin, J. (1988). Multiple Group Factor Analysis of Multitrait-Multimethod Matrices. *Multivariate Behavioral Research*, 23, 469–479. [http://dx.doi.org/10.1207/s15327906mbr2304\\_4](http://dx.doi.org/10.1207/s15327906mbr2304_4)
- Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., & Velasquez, R. (2004). Psychological testing on the Internet: New problems, old issues. *American Psychologist*, 59, 150–162. <http://dx.doi.org/10.1037/0003-066X.59.3.150>
- Newberry, M., & Birtchnell, J. (2011). Negative relating and offense type. *Journal of Criminal Psychology*, 1, 24–35. <http://dx.doi.org/10.1108/20093829201100003>
- Pasveer, K. A., & Ellard, J. H. (1998). The making of a personality inventory: Help from the WWW. *Behavior Research Methods, Instruments, & Computers*, 30, 309–313. <http://dx.doi.org/10.3758/BF03200659>
- Piedmont, R. L., & Chae, J. H. (1997). Cross-cultural generalizability of the five-factor model of personality: Development and validation of the NEO PI-R for Koreans. *Journal of Cross-Cultural Psychology*, 28, 131–115. <http://dx.doi.org/10.1177/0022022197282001>
- Reips, U.-D. (2000). The web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychology experiments on the Internet* (pp. 89–117). San Diego, CA: Academic Press. <http://dx.doi.org/10.1016/B978-012099980-4/50005-8>
- Schonemann, P. H. (1972). An algebraic solution for a class of subjective metrics models. *Psychometrika*, 37, 441–451. <http://dx.doi.org/10.1007/BF02291220>
- Schonemann, P. H., James, W. L., & Carter, F. S. (1978). COSPA: Common space analysis: A program for fitting and testing Horan's subjective metrics model. *Journal of Marketing Research*, 15, 268–272.
- Schwartz, S. H. (2004). Mapping and interpreting cultural differences around the world. In H. Vinken, J. Soeters, & P. Ester (Eds.), *Comparing cultures: Dimensions of culture in a comparative perspective* (pp. 43–73). Leiden, NL: Brill Academic Publishers.
- Stanton, J. M. (1998). An empirical assessment of data collection using the internet. *Personnel Psychology*, 51, 709–725. <http://dx.doi.org/10.1111/j.1744-6570.1998.tb00259.x>
- The International Test Commission. (2006). International guidelines on computer-based and Internet-delivered testing. *International Journal of Testing*, 6, 143–171. [http://dx.doi.org/10.1207/s15327574ijt0602\\_4](http://dx.doi.org/10.1207/s15327574ijt0602_4)
- Triandis, H. C. (1994). *Culture and social behavior*. New York: McGraw-Hill.
- Valkenburg, P. M., & Peter, J. (2009). Social consequences of the internet for adolescents: A decade of research. *Current Directions in Psychological Science*, 18, 1–5. <http://dx.doi.org/10.1111/j.1467-8721.2009.01595.x>
- Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Newbury Park, CA: Sage.
- Van de Vijver, F. J. R., & Tanzer, N. K. (1997). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology*, 47, 263–279.

Received December 16, 2013

Revision received October 1, 2014

Accepted October 8, 2014 ■

Copyright of Psychological Assessment is the property of American Psychological Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.