Personality and Individual Differences 103 (2016) 53–60

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Cross-cultural measurement invariance of the Eysenck Personality Questionnaire across 33 countries

S.C. Bowden [a,*], D.H. Saklofske [b], F.J.R. van de Vijver [c], N.J. Sudarshan [a], S.B.G. Eysenck [d]

[a] University of Melbourne, Australia
[b] University of Western Ontario, Canada
[c] Tilburg University, The Netherlands
[d] The Institute of Psychiatry, England, United Kingdom

ARTICLE INFO

Article history:
Received 26 October 2015
Received in revised form 18 January 2016
Accepted 6 April 2016

Keywords:
Eysenck Personality Questionnaire
Measurement invariance
Factor analysis
Exploratory structural equation modeling
Cross-cultural comparison

ABSTRACT

Cross-cultural invariance of personality measurement provides important information regarding the universality of personality traits. With the recent release of historical data from 33 countries on the Eysenck Personality Questionnaire (EPQ) the opportunity arose to test the invariance of the three personality dimensions measured by the EPQ, together with the response set scale. Although the factor structure of the EPQ has been much studied in previous decades, there was a need to validate the previously reported four-factor structure using modern factor analytic techniques. As anticipated, both exploratory and confirmatory factor analysis provided unsatisfactory models, for different reasons. Instead, exploratory structural equation modeling provided confirmation of the hypothesized oblique, four-factor structure. Results of measurement invariance comparisons across countries, separately by gender, showed striking evidence of the generalizability of the four-factor structure in pairwise comparisons with the English sample as the reference group. In addition, there was evidence of failure of item invariance, for a varying subset of items, in all of the between-country comparisons. However, the pattern of partial measurement invariance does not preclude effective use of the EPQ as a research tool in diverse cultures. The results provide strong theoretical convergence, with other published studies, on the universality of the four-factor structure.

Following from the Eysenck Personality Inventory, the Eysenck Personality Questionnaire (EPQ; (Eysenck & Eysenck, 1975) was introduced as a measure of three primary dimensions of personality. Eysenck's model defined the three dimensions or factors as neuroticism (N), extraversion (E) and psychoticism (P), the latter sometimes described as tough-mindedness (Eysenck & Eysenck, 1975; van Hemert, van de Vijver, Poortinga, & Georgas, 2002). In addition, the EPQ contains a social-desirability (L or Lie) scale designed to measure a symptom-minimization response set. Eysenck and Eysenck (1975) argued that these three primary dimensions of personality (N, E and P) captured most of the variance in more elaborate models of personality and therefore provided an efficient measure of individual differences. Although supported by a wide variety of construct validity evidence, the factor structure of the EPQ has been controversial. Some studies have reported the predicted factor-structure whereas other studies have failed to identify or replicate the three factor-structure (for reviews see Barrett & Kline, 1980; Barrett & Kline, 1982; Roger & Morris, 1991). Cross-cultural replication of a factor structure is often taken as the litmus test for the universality of a personality structure, if an instrument yields invariably the same factor structure across a wide range of cultures, the personality structure captured by the instrument is taken to be universal (van de Vijver & Leung, 1997). In this paper we review earlier ways to statistically evaluate the evidence for invariance and then present and apply recent developments in factor-analytic procedures to the extensive cross-country EPQ data sets. Thus, we address the global applicability of Eysenck's three factor structure combined with the single response-set scale (L).

Factor-analytic studies of the EPQ, conducted over many years, have used what are now recognized as suboptimal methods including earlier approaches to Exploratory Factor Analysis (EFA) and Principal Component Analysis (PCA). Although widely used to report factor structures, PCA does not belong to the family of common-factor methods, namely, methods that distinguish common variance, attributable to the hypothesized factor structure, from unique variance in observed scores (Preacher & MacCallum, 2003). Therefore, PCA is best avoided if factor analysis is the intent (Floyd & Widaman, 1995; Henson & Roberts, 2006). In addition, older approaches to EFA rely less on goodness-of-fit criteria, instead

* Corresponding author at: Melbourne School of Psychological Sciences, University of Melbourne, Parkville, Victoria 3010, Australia.
E-mail address: sbowden@unimelb.edu.au (S.C. Bowden).

http://dx.doi.org/10.1016/j.paid.2016.04.028
0191-8869/

employing various arbitrary rules to select the number of factors (Floyd & Widaman, 1995; Henson & Roberts, 2006). These older EFA techniques are prone to sample-specific solutions that do not replicate well, especially when the factor analysis is based on item-level data with lower communalities (Floyd & Widaman, 1995).

As a consequence, it has been recommended for many years that if a hypothesized factor structure is known, then EFA in general should be avoided, and instead, Confirmatory Factor Analysis (CFA) should be the method of choice to replicate and improve understanding of latent structures (Floyd & Widaman, 1995; Henson & Roberts, 2006). However, any brief survey of the contemporary factor analytic literature suggests that these recommendations are commonly ignored, with consequent proliferation of poorly replicated factor solutions for many cognitive, personality and psychopathology tests.

Importantly, most current versions of factor-analytic software provide improved goodness-of-fit information for EFA as well as for CFA output, however, analysis of the EPQ to date has not taken advantage of these developments. In addition, the factor structure of many personality and other inventories, such as Big-Five inventories, has also been controversial, because of the difficulty establishing a well-fitting CFA model. Examination of replicability across populations is best undertaken with multiple-group CFA which can examine the precise numerical replicability of factor models across populations (Meredith, 1993; Widaman & Reise, 1997). Recent advances have extended the notion of invariance focusing on approximate invariance, which relaxes the requirement of invariance of all parameters across all groups studied. Thus, in Bayesian Structural Equation Modeling, parameters (notably factor loadings and intercepts) are allowed to vary with an a priori defined variance (Muthén & Asparouhov, 2012). In the so-called alignment method it is possible to identify subgroups of countries with invariant parameters (Asparouhov & Muthén, 2014). Both approaches stay within the CFA framework and deal with the often observed poor fit of scalar invariance models by examining related, yet more flexible types of invariance. Another way of dealing with fit problems in comparisons involving many groups is another extension of CFA, namely Exploratory Structural-Equation Modeling (ESEM: Asparouhov & Muthén, 2009; Marsh et al., 2010). The ESEM approach, used in the present study, does not require the restrictive assumptions of CFA simple-structure or the a priori specification of Bayesian approaches. In particular, ESEM avoids the theoretically tenuous assumption that items in personality tests are single-factor or trait-specific (Marsh et al., 2010). The ESEM approach also provides better articulation of oblique or correlated factor-structures and avoids the problems associated with potentially biased estimates of factor covariance, biased upward because of the restrictive simple-structure assumptions of CFA. Biased estimates of personality factor covariances may further bias convergent and discriminant validity studies (Marsh, Morin, Parker, & Kaur, 2014; Marsh et al., 2010).

The issues associated with different generations of factor-analytic methods are well illustrated in the earlier work of Barrett and colleagues (Barrett, Petrides, Eysenck, & Eysenck, 1998) who examined the generalization of the four-factor model of the 90-item EPQ across 34 countries. Barrett and colleagues employed variations on PCA methods to assess the geometric generalization of rotated PCA solutions. In their earlier study they used the so-called KHB method to test the geometric precision of PCA solutions across countries (Barrett & Eysenck, 1984). They argued in favor of the universality of Eysenck's three-factor model of personality. However, the KHB method was subject to criticism (Bijnen, van der Net, & Poortinga, 1986), so in a later study Barrett et al. (1998) examined alternative congruence coefficients to test the similarity of PCA solutions across 34 countries, concluding that the congruence coefficients provided support for the similarity of factor solutions across countries. However, as noted above PCA solutions are now regarded as suboptimal. In addition, Barrett and colleagues were publishing their approach on the cusp of wide availability of multiple-group mean-structure CFA algorithms (see Widaman & Reise, 1997).

With the recent public release of extensive EPQ data sets, the opportunity arises to re-examine the factor structure of EPQ using rigorous, contemporary factor-analytic methodology. As in the case of recent re-analysis of Big-Five data sets, the application of contemporary factor-analytic methods may aid theoretical refinement. CFA and by extension ESEM provide the strongest available techniques to examine convergent and discriminant validity (Strauss & Smith, 2009).

In addition, the variety of historical data from different countries allows replication and extension of the, as yet, limited data on cross-cultural generalization of personality traits. Some cross-cultural invariance research exists in relation to Big-Five scales and other specific personality measures (Marsh et al., 2010). However, the extensive EPQ data sets provide a unique opportunity to examine the invariance of these three major personality dimensions together with the Lie scale across many countries.

Examination of measurement invariance involves the evaluation of the precise numerical generalization of the latent-factor model relating observed item scores to factors across populations (Brown, 2015; Meredith & Teresi, 2006; Widaman & Reise, 1997). Establishing measurement invariance is regarded as an essential precursor to meaningful construct validity research in its broadest sense (Bowden, Saklofske, & Weiss, 2011; Meredith & Teresi, 2006; Widaman & Reise, 1997). With continuous indicators, establishing measurement invariance is algebraically equivalent to demonstrating that the regression relationship between any observed item score and the respective factor score in the CFA model is identical across populations. With categorical indicators typically used in personality inventories including the EPQ, establishing measurement invariance is algebraically equivalent to demonstrating that the item-response curve for any item is identical across populations (Bandalos, 2008; Millsap & Yun-Tein, 2004). Establishing measurement invariance across populations permits a wide variety of desirable construct validity inferences including that the observed test scores convey the same psychological meaning in the respective populations.

It has been argued that gender is a potentially confounding variable in cross-national comparisons of personality and that gender differences influence internal consistencies, means, and factor structures. Miles, Shevlin and McGhee (1999) found no gender differences in internal consistencies in the EPQ-R in a British sample of patients with skin diseases. However, there is considerable evidence for gender differences in personality. Costa, Terracciano and McCrae (2001) working with Revised NEO Personality Inventory, found that across 26 cultures females reported themselves to be higher in neuroticism, agreeableness, warmth, and openness to feelings, and lower in assertiveness and openness to ideas than males. Moreover, these gender differences were larger in Western countries than in non-Western countries, whereas Feingold (1994) found consistent gender differences across nations in a meta-analysis covering the period 1940–1992. Martin and Kirkcaldy (1998) found in a study of 100 students from Northern Ireland that females scored higher in neuroticism and lower in psychoticism than males, while Lynn and Martin (1997) analyzed data from 37 countries and found that females scored higher on neuroticism in all countries and lower on psychoticism and extraversion in most countries. Gender differences were unrelated to per capita incomes. Despite the consistent findings of gender differences in mean scores, there is support for the generality of factor structure of personality measures across dichotomous gender categories (e.g., McCrae & Costa, 1997).

The aims of this study were to validate the four-factor model of EPQ scores (three personality dimensions and the lie scale as a measure of dissimulation or social desirability) in the historical data sets from a heterogeneous sample of 33 countries, test the preferred model using contemporary factor estimation techniques and then test for measurement invariance across national samples. Furthermore gender will also be examined as normative and research studies have suggested at least small mean (and variance) differences in personality across gender which warrants another look at the invariance of the factor structure in males and females.