Personality and Individual Differences 51 (2011) 829–835

Contents lists available at ScienceDirect

Personality and Individual Differences

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

Psychometric analysis of the Empathy Quotient (EQ)

C. Allison [a,*], S. Baron-Cohen [a], S.J. Wheelwright [a], M.H. Stone [b], S.J. Muncer [c]

[a] Autism Research Centre, Department of Psychiatry, University of Cambridge, Douglas House, 18B Trumpington Road, Cambridge CB2 8AH, UK
[b] Psychology and Mathematics Departments, Aurora University, 347 South Gladstone Avenue, Aurora, IL 60506-4892, USA
[c] Clinical Psychology, Teesside University, Middlesbrough, TS1 3BA, UK

ARTICLE INFO

Article history:
Received 4 November 2010
Received in revised form 30 June 2011
Accepted 4 July 2011
Available online 6 August 2011

Keywords:
Autism spectrum conditions
Empathy
Rasch
Confirmatory Factor Analysis

ABSTRACT

This study assessed the dimensionality of the Empathy Quotient (EQ) using two statistical approaches: Rasch and Confirmatory Factor Analysis (CFA). Participants included $N = 658$ with an autism spectrum condition diagnosis (ASC), $N = 1375$ family members of this group, and $N = 3344$ typical controls. Data were applied to the Rasch model (Rating Scale) using WINSTEPS. The Rasch model explained 83% of the variance. Reliability estimates were greater than .90. Analysis of differential item functioning (DIF) demonstrated item invariance between the sexes. Principal Components Analysis (PCA) of the residual factor showed separation into Agree and Disagree response subgroups. CFA suggested that 26-item model with response factors had the best fit statistics (RMSEA.05, CFI .93). A shorter 15-item three-factor model had an omega (ω) of .779, suggesting a hierarchical factor of empathy underlies these sub-factors. The EQ is an appropriate measure of the construct of empathy and can be measured along a single dimension.

1. Introduction

Empathy allows us to make sense of the behaviour of others, predict what they might do next, how they feel and also feel connected to that other person, and respond appropriately to them (Wheelwright & Baron-Cohen, 2011). Empathy involves an affective and a cognitive component (Baron-Cohen & Wheelwright, 2004). The former relates to an individual having an appropriate emotional response to the mental state of another. The latter largely overlaps with the concepts of 'mindreading', or 'theory of mind': the ability to attribute mental states to others; an understanding that other people have thoughts and feelings, and that these may not be the same as your own (Baron-Cohen, 1995). Baron-Cohen and Wheelwright (2004) argue that these two components of empathy co-occur and cannot be easily disentangled.

The Empathy Quotient (EQ) (Baron-Cohen & Wheelwright, 2004) was developed as a measure of empathy because of shortcomings in existing instruments like the Interpersonal Reactivity Index (IRI) (Davis, 1980), the Questionnaire Measure of Emotional Empathy (QMEE) (Mehrabian & Epstein, 1972) and the Empathy (EM) Scale (Hogan, 1969) (see Baron-Cohen & Wheelwright, 2004; Lawrence, Shaw, Baker, Baron-Cohen, & David, 2004). The EQ is sensitive to differences in empathy in clinical and general populations; individuals with an autism spectrum condition (ASC) have reduced levels of self-reported empathy (measured by the EQ), relative to typical controls (Baron-Cohen & Wheelwright, 2004; Berthoz, Wessa, Kedia, Wicker, & Grezes, 2008; Kim & Lee, 2010; Lawrence et al., 2004; Wakabayashi et al., 2007; Wheelwright et al., 2006). The EQ shows a sex difference in empathy in the general population, females on average having higher scores than males (Baron-Cohen & Wheelwright, 2004). These findings have been replicated in cross-cultural studies in Japan (Wakabayashi et al., 2007), France (Berthoz et al., 2008) and Italy (Preti et al., 2011). A study in Korea (Kim & Lee, 2010) did not find an overall sex difference in total EQ score, an anomaly that needs to be tested further. A child parent-report version of the EQ showed a similar pattern of sex differences to that observed in adults (Auyeung et al., 2009). The EQ has clinical utility and is used as part of a screening protocol along with the Autism Spectrum Quotient (AQ) (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) for a clinical assessment in an adult diagnostic clinic for ASC (Baron-Cohen, Wheelwright, Robinson, & Woodbury-Smith, 2005). The EQ has convergent validity; it correlates with the 'Reading the Mind in the Eyes' Test (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001) and the Toronto Alexithymia Scale (TAS) (Lombardo et al., 2009). The EQ has been found to inversely correlate with foetal testosterone (FT) levels (Chapman et al., 2006), with single nucleotide polymorphisms (SNPs) in genes related to sex steroid hormones, neural growth, and social reward (Chakrabarti et al., 2009), and with neural activity during emotion perception in fMRI (Chakrabarti, Bullmore, & Baron-Cohen, 2006).

Lawrence et al. (2004) examined the factor structure of the EQ using Principal Components Analysis (PCA) and found a three factor solution (consisting of cognitive empathy, emotional reactivity and social skills). Berthoz et al. (2008) confirmed this structure using Confirmatory Factor Analysis (CFA). Muncer and Ling

* Corresponding author. Tel.: +44 1223 746082; fax: +44 1223 746033.
E-mail address: cla29@cam.ac.uk (C. Allison).

0191-8869/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.paid.2011.07.005

(2006) tested the unidimensionality of the EQ using CFA and found this model did not adequately fit their data. They tested other structures and confirmed that a three factor solution consisting of 15 items best fit the data. Kim and Lee (2010) confirmed this structure in their Korean sample. To date, investigation of the dimensionality of the EQ has been limited to the application of factor analysis (FA). Differences in response options (agreeing or disagreeing to selected items) may give rise to finding factors that are potentially absent or theoretically meaningless even after reverse coding of items towards the appropriate direction has occurred. Lawrence et al. (2004) point out that factor analysis on ordinal data can result in spurious factors where items load according to 'difficulty' (Gorsuch, 1974). To this end, we take a different approach at examining the dimensionality of the EQ using Rasch analysis in combination with CFA.

1.1. The Rasch model

In classical test theory (CTT), ordinal responses to questionnaire items are often treated as interval. This can lead to erroneous conclusions and inferences about the scale especially when a sum score is used to define the degree to which an individual possesses a trait or characteristic (Santor & Ramsay, 1998). Rasch (1960) developed a unique approach to psychometrics which fulfils the requirements of additive measurement (Perline, Wright, & Wainer, 1979). The principle behind Rasch analysis is as follows: 'A person having a greater ability than another person should have the greater probability of solving any item of the type in question, and similarly, one item being more difficult than another means that for any person the probability of solving the second item is the greater one' (Rasch, 1960). When participants complete a psychometric scale they provide two sources of information. One informs us how people respond to the items, (used in reliability and factor analysis studies), and the other how the participants score on the scale. This latter information is not much used in CTT. Rasch's approach uses both pieces of information when scales are analysed. The probabilistic relationship is modelled between person ability and item difficulty as a latent trait. It locates person ability and item difficulty along the same continuum in logits or log odds. The Rasch model transforms data from ordinal scores into interval level measurement with the logit.

Item difficulty is calculated using the proportion of participants who get the answer 'correct'. This is transformed into the log odds probability of getting the item correct. The ability of each participant can also be calculated, by taking the percentage of items they get correct and turning this into a probability of answering an item correctly. Rasch's theory suggests that the probability of getting an individual item correct is produced by the difference between a person's ability and the item difficulty. If a person's ability is higher than an item's difficulty, then the participant is more likely to get this correct than if it is lower than the item's difficulty. Using this information the data collected can be compared with what would be expected based on calculations of item difficulty and person ability. The closer the results are to the predicted results, the better fit the data are to the Rasch model.

Rasch analysis is designed to produce unidimensional measures when the data fit the model. Therefore, the instrument measures only one ability/personality trait/attitude. It is also designed to produce measures in which the difference between participant scores is interval scaled, making it more appropriate for statistical analysis. Rasch analysis satisfies the criteria for simultaneous conjoint measurement (Karabatsos, 2001). If the measure is unidimensional then it is reasonable to sum the item scores to produce a total score that is an adequate representation of the measured dimension. The count must be of a cohesive unit otherwise the count/measure is invalid. Rasch analysis will transform the raw counts into these cohesive units while CFA analyses the qualities of the raw ordinal (rather than interval) counts. From a Rasch perspective, items are selected to cover a wide range of the dimension, while CFA includes items that maximise reliability. Further, Rasch measures are less sensitive to directional factors (Singh, 2004) than are CFA measures.

The Rasch model has been criticised recently as not being an example of conjoint measurement (Kyngdon, 2008) (although see Michell (2008) for criticisms of Kyngdon's argument). Rasch analysis emphasises producing unidimensional measures; the main purpose of the EQ is to provide a reliable and valid measure of empathy. However, CFA is regarded as one of the most important methods for examining psychometric properties.

1.2. Aims and Objectives

We will take a pragmatic approach to examining the dimensionality of the EQ. The aims are to apply the Rasch model to a large EQ dataset to create a unidimensional measure of empathy. We then examine this model and other proposed EQ models using CFA.

2. Methods

2.1. Data source

Data included in the analysis were collected at the websites of the Autism Research Centre (ARC), University of Cambridge. Individuals can register as research volunteers and complete online questionnaires and tests. The ARC website (www.autismresearchcentre.com) recruits individuals with ASC as well as parents of children with ASC. Individuals from the general population who have an interest in taking part in research can register at www.cambridgepsychology.com. Everyone is invited to complete the Empathy Quotient (EQ). Altogether 5377 individuals completed the EQ online of which 3265 were female and 2112 were male. Within this sample, 658 individuals had a diagnosis of ASC, 1375 were family members of an individual with ASC, and 3344 had no diagnosis of ASC. The mean age of the whole sample was 30.4 years (SD = 11.4, range 16.0–78.0).

2.2. The EQ

The EQ consists of 40 statements to which participants have to indicate the degree to which they agree or disagree. There are four response options: 'strongly agree', 'slightly agree', 'slightly disagree', 'strongly disagree'. 'Definitely agree' responses score two points and 'slightly agree' responses score one point on half the items, and 'definitely disagree' responses score two points and 'slightly disagree' responses score one point on the other half. The remainder of the response options score 0. See Baron-Cohen and Wheelwright (2004) for full details.

2.3. Rasch analysis

Rasch analysis was conducted using the Rating Scale (Andersen, 1977) routine in WINSTEPS (Linacre, 2006). PROX estimation was used to converge the data with the Rasch model. The WINSTEPS reliability estimate was executed to provide an estimate of cohesion of the items (in terms of person and item reliability estimates). Item and person misfit and item Infit and Outfit statistics were examined.

Point-biserial correlations between items scores and total score were examined. It is generally agreed that these coefficient values are most acceptable for item discrimination when they occur between 0.2 and 0.8, or even closer between 0.3 and 0.7. Hence,