

Measuring empathy: reliability and validity of the Empathy Quotient

E. J. LAWRENCE, P. SHAW, D. BAKER, S. BARON-COHEN AND A. S. DAVID*

Section of Cognitive Neuropsychiatry, Department of Psychological Medicine, Institute of Psychiatry, DeCrespigny Park, Denmark Hill, London SE5 8AF, UK; Autism Research Centre, University of Cambridge, Departments of Experimental Psychology and Psychiatry, Downing Street, Cambridge CB2 3EB, UK

ABSTRACT

Background. Empathy plays a key role in social understanding, but its empirical measurement has proved difficult. The Empathy Quotient (EQ) is a self-report scale designed to do just that. This series of four studies examined the reliability and validity of the EQ and determined its factor structure.

Method. In Study 1, 53 people completed the EQ, Social Desirability Scale (SDS) and a non-verbal mental state inference test, the Eyes Task. In Study 2, a principal components analysis (PCA) was conducted on data from 110 healthy individuals and 62 people reporting depersonalisation (DPD). Approximately 1 year later, Study 3, involved the re-administration of the EQ ($n=24$) along with the Interpersonal Reactivity Index (IRI; $n=28$). In the last study, the EQ scores of those with DPD, a condition that includes a subjective lack of empathy, were examined in depth.

Results. An association was found between the Eyes task and EQ, and only three EQ items correlated with the SDS. PCA revealed three factors: (1) ‘cognitive empathy’; (2) ‘emotional reactivity’, and (3) ‘social skills’. Test–retest reliability was good and moderate associations were found between the EQ and IRI subscales, suggesting concurrent validity. People with DPD did not show a global empathy deficit, but reported less social competence.

Conclusions. The EQ is a valid, reliable scale and the different subscales may have clinical applications.

INTRODUCTION

There are several definitions of empathy reflecting its multidimensional nature. Social psychologists have conceptualized empathy as having two main strands (1) cognitive empathy – ‘the intellectual/imaginative apprehension of another’s mental state’ and (2) emotional empathy – ‘an emotional response to ... emotional responses of others’. Recently, in the literature, emotional empathy has also been labelled ‘affective’ empathy. The literature on ‘theory of mind’ (or the ability to think about the contents

of other minds) overlaps with cognitive empathy and the terms are used interchangeably here.

For an emotional response to count as ‘affective empathy’ it has to be appropriate to the observed mental state. Emotional responses to others’ mental states can be classified as: (1) parallel – the response matches that of the target, for instance, feeling fear at another’s fright, and (2) reactive – involves going beyond a simple matching of affect – such as sympathy or compassion (Davis, 1994). However, some emotional responses are not considered truly empathic, i.e. happiness at another’s misfortune or, less obviously, ‘personal distress’ (Davis, 1980; Eisenberg *et al.* 1987). The latter occurs when someone has a self-orientated state of ‘personal distress’ in response to another’s

* Address for correspondence: Professor Anthony David, Section of Cognitive Neuropsychiatry, Box 68, Institute of Psychiatry, DeCrespigny Park, London, SE5 8AF, UK.
(Email: a.david@iop.kcl.ac.uk)

negative state (Batson *et al.* 1987). What distinguishes this from an empathic response is that it is self- rather than other-orientated.

Several scales have been developed to measure empathy but each has important weaknesses. The Questionnaire Measure of Emotional Empathy (Mehrabian & Epstein, 1972) was designed to tap emotional empathy. However, with hindsight, the authors suggest it may measure general emotional arousability instead (Mehrabian *et al.* 1988). Items on a newer version – the Balanced Emotional Empathy Scale (Mehrabian, 2000) – measure, more specifically, reactions to others' mental states, but unfortunately, it is still not clear that they tap emotional empathy alone, e.g. 'I cannot easily empathise with the hopes and aspirations of strangers/I easily get carried away by the lyrics of a love song'. A questionnaire measuring cognitive empathy (Hogan, 1969) was also developed in the 1960s; however, a factor analysis suggested it may actually tap social self-confidence, even temperedness, sensitivity and non-conformity (Johnson *et al.* 1983). Critics also argue that it measures simply social skills rather than empathy *per se* (Davis, 1994).

The Interpersonal Reactivity Scale (Davis, 1980) adds further dimensions to the measurement of empathy. It includes subscales that measure perspective-taking, in line with traditional definitions of cognitive empathy, empathic concern which specifically addresses the capacity of the respondent for warm, concerned, compassionate feelings for others, fantasy items – which measure a tendency to identify with fictional characters and personal distress which is designed to tap the occurrence of self-orientated responses to others' negative experiences. The author describes the questionnaire as tapping four separate aspects of empathy but it is unclear whether the fantasy subscale taps pure empathy (Baron-Cohen & Wheelwright, in press) – and personal distress, despite being important, is not empathy in itself.

The EQ (Baron-Cohen & Wheelwright, in press) (see website for Appendix 1) is the most recent addition, and unlike previous scales it was explicitly designed to have a clinical application and be sensitive to a lack of empathy as a feature of psychopathology. Several groups

have been hypothesised as having problems employing 'empathy'. Most obvious, are those diagnosed with autistic spectrum disorders and people who display signs of psychopathy (Blair, 1995). More recently, other groups have been suggested, such as those who report depersonalisation (Senior *et al.* 2001; Baker *et al.* 2003), who frequently complain of a subjective deficit in empathising.

The EQ was validated on 197 healthy control volunteers and 90 people with Asperger's Syndrome and High-functioning Autism (AS/HFA) and age and sex matched controls (a sex ratio of 2.6:1 m:f was found). It was shown to distinguish reliably between the clinical and control groups. The authors also found sex differences in the control group with women scoring significantly higher. In addition, the EQ was found to have high test-retest reliability over a period of 12 months. Baron-Cohen *et al.* (2003) replicated the female superiority on the EQ and showed once again that it distinguished between those with AS/HFA and controls.

The aim of this paper was to examine further the validity and reliability of the EQ across samples. Test-retest reliability was re-examined, and the association between the EQ and a well-validated measure of 'social desirability' (Crowne & Marlowe, 1960) was explored. This was included to address a general problem with self-report measures, that is that people may respond according to how they would like to appear, i.e. highly 'empathic'. The association between the EQ and the Eyes task (Baron-Cohen *et al.* 2001) was also considered as a means of assessing construct validity. Next, an exploratory factor analysis was performed in order to explore the various components of empathy. As a further check on concurrent validity, the relationship between the EQ and the Interpersonal Reactivity Index (IRI; Davis, 1980) was then examined. Lastly, the EQ scores of people with DPD were considered in depth.

Study 1

Participants

There were 53 volunteers [28 (52.8%) women and 25 (47.2) men] with a mean age of 32.5 years (± 10.9). Approximately, 50% of this group were recruited from mental health professionals at the Institute of Psychiatry (40% of men and

60% of women). The remainder were recruited from non-academic/clinical staff and through advertisements in the local area.

Procedure

All measures were completed in a quiet room as part of a wider testing session. Participants were given the EQ (Baron-Cohen & Wheelwright, in press) self-report measure of empathy. Responses are given on a 4-point scale ranging from 'strongly agree' to 'strongly disagree'. Approximately half the items are reversed. Participants received 0 for a 'non-empathic' response, whatever the magnitude, and 1 or 2 for an 'empathic response' depending on the strength of the reply. There are 60 items including 20 filler items – and so the total score is out of 80. Missing values on the EQ, resulting from a double endorsement or no endorsement, were substituted with the group mean rounded to the nearest whole number.

Participants were also given the Social Desirability Scale (SDS; Crowne & Marlowe, 1960) which taps people's tendency to respond to items in a socially desirable way. One point is allocated for each item endorsed, resulting in scores ranging from 1 to 33 with a high score indicating that the respondents are prone to give answers which show themselves in a good light, i.e. 'I sometimes feel resentful when I don't get my own way'.

The Eyes test (Baron-Cohen *et al.* 2001; Shaw *et al.* 2003) was also administered. This measures peoples' ability to decipher a mental state from pictures of the eyes alone and according to the authors, is an advanced measure of mind-reading or in our terminology 'cognitive empathy'. This test has been shown to distinguish reliably between people with AS/HFA and healthy individuals. One point is allocated for each correct answer with a final score out of 36.

Lastly, participants completed the National Adult Reading Test (Nelson, 1982). Participants read 50 irregular sounding words (i.e. ache), which yields an estimate of IQ.

Results

Mean total EQ scores for both men and women can be found in Table 1. These are similar to those found in the original study (Baron-Cohen & Wheelwright, in press), i.e. males 41.8 (± 11.2)

Table 1. Mean and s.d. scores on the EQ

	n	Total score on the EQ			
		Mean	s.d.	Min	Max
Male	25	41.3	10.1	22	58
Female	28	50.6	9.2	30	66
Group total	53	46.2	10.6	22	66

and females 47.2 (± 10.2). Sex differences were also found ($t = -3.5$, $df = 51$, $p = 0.001$). The data were normally distributed [slightly negative skew (-0.190) and kurtosis of less than 1 (-0.717)].

Each item on the EQ was entered into a Pearson's Product Moment Correlation analysis along with the total score on the SDS. A positive correlation above 0.3 was taken as an indicator of socially desirable responding. Items 11, 18, 27, 34 and 37 of the EQ, all correlated significantly with total SDS score but item 27 correlated below 0.3, and item 37 had a negative rather than positive relationship. Items 11, 18 and 34 were therefore dropped from subsequent analyses.

The mean score on the Eyes test was 27.6 (± 4) which is very similar to the normative data (general population 26.2/students 28). These data were then correlated with total EQ score and a modest positive relationship was found ($n = 48$, $r = 0.294$, $p = 0.033$).

The estimated IQ score from the NART for this group was 120.48 (± 4.7) which is above the average range. As both the Eyes test and EQ have verbal components, a correlational analysis was run to examine the association between verbal IQ, as estimated from the NART, and each of these variables. There was a near significant association between performance on the Eyes test and verbal IQ ($n = 48$, $r = 0.385$, $p = 0.07$) but not between the total EQ score and verbal IQ.

A multiple regression analysis was performed to include total EQ score, verbal IQ and other demographic factors (sex, age, education and whether the participant was a clinician/academic or not). The only significant predictor of the Eyes test was verbal IQ score (multiple $r = 0.369$) which accounted for 11.7% of the variance. However, both sex ($r = 0.266$, $t = 1.83$, $p = 0.074$) and EQ score ($r = 0.255$, $t = 1.75$, $p = 0.087$) also approached significance.

Study 2

Participants

An additional 57 volunteers [22 men (38.6%) and 35 (61.4%) women] completed the EQ. These participants were recruited by the first two authors during the course of other projects. These data were combined with those from Study 1 to create a control group of 110 psychologically healthy participants.

In addition, 54 people who contacted the Depersonalization Research Unit at the Institute of Psychiatry, London, reporting symptoms of depersonalization disorder (DPD), were sent the EQ along with some initial mental health screening measures. Some of these people are a subgroup of a cohort reported elsewhere (Baker et al. 2003). A further eight people diagnosed with (DPD) at the same unit were also recruited. They completed the EQ during an experimental testing session along with other cognitive measures. As a whole, this group comprised 32 men (51.6%) and 30 women (48.4%), with a mean age of 34.6 (± 10.8).

DPD is defined as an 'alteration in the perception or experience of the self so that one feels detached from and as if one is an outside observer of one's mental processes or body' (DSM-IV, 1994). People with DPD also often report a lack of subjective empathy, although the cause and nature of this is unclear. Despite this, there is no reason to expect any difference between the EQ factor structures between the DPD group and healthy individuals, although there may well be a difference in scores.

A χ^2 analysis revealed that the gender distribution was not significantly different between the control and the DPD groups ($\chi^2=1.26$, $df=1$, $p>0.05$). Neither were ages significantly varied between these two groups ($t=-0.593$, $df=113$, $p>0.05$). For the purposes of analysis, all the groups were combined resulting in 79 (45.9%) men and 93 (54.1%) women [mean age 34.1 years (± 10.4)].

Procedure

An exploratory factor analysis, using a principal components analysis (PCA) to construct the initial model, was performed on the EQ. Although the data are ordinal, many authors feel that this procedure is still useful as long as meaningful factors are extracted (Hutcheson & Sofroniou,

Table 2. Mean and s.d. EQ scores for entire sample

	n	Total score on the EQ			
		Mean	s.d.	Min	Max
Male	79	40.9	11.9	15	66
Female	93	49.6	9.6	23	69
Group total	172	45.6	11.6	15	69

1999). The main worry is that it can result in spurious factors where items load according to 'difficulty' (Gorsuch, 1974) and/or that the factors may be harder to interpret (Kim & Mueller, 1978). Both of these issues were kept in mind when interpreting the analysis.

Nine cases had missing values ranging from 1 to 4 and were dealt with as described in Study 1. However, one additional participant had a whole page missing and so these values were left as missing.

Results

The mean EQ scores (see Table 2) are remarkably similar to the normative data for both men (mean 41.8 ± 11.2) and women (47.2 ± 10.2) including sex differences ($t=-5.34$, $df=147.38$, $p=0.001$).

Group comparison

A separate analysis was conducted for each group (DPD v. healthy volunteers) to examine the similarity of the factor structure. A PCA followed by an exploratory factor analysis was performed with a varimax rotation. Scree plots were used (Cattell, 1966), as opposed to eigen values which can give rise to many uninterpretable factors. Values less than 0.3 were suppressed.

A salient loading profile (Abdel-Khalek et al. 2002) was performed using 0.35 as a cut-off point (see Table 3). These figures were considered along with tentative labels for each of the factors (Tabachnick & Fidell, 1989) and the decision was made to combine the data.

Data screening

A Pearson's correlation matrix was generated and all EQ items that failed to correlate with any other item at 0.2 (Hutcheson, 1999) or had low communalities in the final model, were removed,

Table 3. *Salient loading analysis*

	No. of salient loadings		Common loadings	
	Control group	DPD group	<i>n</i>	%*
Factor 1	12	17	12	100
Factor 2	10	10	8	80
Factor 3	10	9	5	50

* The percentages were calculated in proportion to the control group salient loadings.

namely 15, 18, 28, 37, 38, 39, 49, 60 (see website for Appendix 1).

All EQ items were also re-correlated with the total SDS score. Five EQ items were significantly associated with the total SDS score, namely 11, 18, 34, 37 and 46. Item 37 again showed a negative relationship; however, it also had a low loading as did item 18 (see above), and so this stage of data screening only resulted in the removal a further three items, i.e. 11, 34 and 46. Eleven items were, therefore, left out of the analysis.

Final analysis

There were 29 items and 172 cases, conforming to the five cases per item rule. A PCA with a varimax rotation showed the communalities to lay in the mid range except for items 10 and 57. No. 57 was kept as it loaded onto factor 3 in the final model and no. 10 was removed, as it did not load onto any factors, leaving 28 items in total.

The scree plot showed that only three or four plots (factors) appeared stacked and separate from the rest with the remaining plots falling away and bunched together (see website for Fig. 1). Three factors were kept as it was apparent from both the scree plot and eigen values that they were the strongest, accounting for 41.4% of the total variance.

The item loading for these three factors in the rotated solution are shown in Table 4. Double loadings were allocated on the basis of content, with agreement reached between the first and second authors. The Keiser–Meyer–Olkin measure of sampling adequacy was 0.846 and the Bartlett test of sphericity was highly significant, suggesting the data were suitable for PCA. Factor 1 was labelled ‘cognitive empathy’, factor 2 ‘emotional reactivity’ and factor 3 ‘social skills’.

Table 4. *Final loadings from principal components analysis*

	1	2	3
EQ55	0.763		
EQ52	0.726		
EQ25	0.723		
EQ54	0.696		
EQ44	0.688		
EQ58	0.680		
EQ26	0.658		
EQ41	0.633		
EQ19	0.583		
EQ36	0.559	0.315	
EQ1	0.505		0.315
EQ32		0.675	
EQ59		0.658	
EQ42		0.593	
EQ21		0.528	
EQ48		0.508	
EQ6		0.497	
EQ27		0.473	
EQ50		0.466	
EQ43	0.442	0.452	
EQ22	0.322	0.385	
EQ29		0.333	
EQ8			0.771
EQ35			0.768
EQ12			0.619
EQ14			0.575
EQ4			0.538
EQ57			0.398

Validity

The relationship between factors was explored and factors 1 and 2 correlated significantly ($n=171, r=0.497, p=0.0001$) as did factors 1 and 3 ($n=171, r=0.254, p=0.001$) and 2 and 3 ($n=171, r=0.209, p=0.006$). These associations were as expected; however, the co-efficients are not so high as to preclude discriminant validity.

A 3×2 repeated-measures ANOVA was conducted to examine the sex differences by factor. Factor scores were used, as they are a more accurate index of a person’s score on a particular factor. Again, there was a main effect of sex [$F(1, 169)=19.46, p=0.001$] and an interaction between sex and the scores on the different factors [$F(2, 338)=5.85, p=0.003$]. *t* tests revealed significant differences on ‘cognitive empathy’ ($t=-3.083, df=169, p=0.002$) and on ‘emotional reactivity’ ($t=-4.725, df=169, p=0.001$) but not on ‘social skills’ ($t=0.206, df=169, p>0.05$).

There was a significant correlation between performance on the Eyes task and the factor scores for ‘social skills’ ($n=53, r=0.273,$

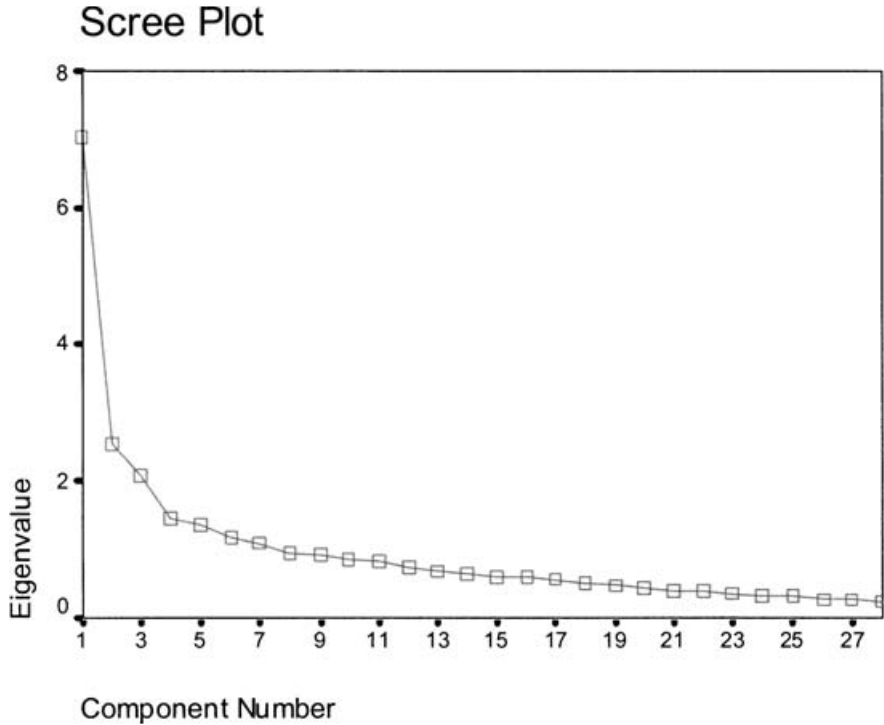


FIG. 1. Scree plot for entire sample.

$p=0.048$), but not with either of the other factors. Each factor score was also correlated with predicted verbal IQ as a further check but none had a significant association. The Eyes score, factor scores and demographics (see Study 1) were then entered into a multiple regression analysis. Again, verbal IQ was the only significant predictor with sex approaching significance (see Study 1 for statistics).

Study 3

Participants

Forty-four people were re-contacted 10–12 months after Study 1. A further 4 people who had not taken part in the first studies also participated. The final group consisted of 29 people [11 males (37.9%) and 18 females (62.1%)] with a mean age 32 years (± 9.5). There were no age differences between this group and the participants in Study 1 ($t=1.29$, $df=51$, $p>0.05$) nor was there any difference in sex distribution ($\chi^2=1.41$, $df=1$, $p>0.05$). Twenty-four of the original participants returned the EQ and IRI,

one person just returned the EQ, and an additional four people filled out the IRI and EQ at time 2 only.

Procedure

Participants were sent both the EQ and the IRI (Davis, 1980). The EQ was re-sent in order to replicate the test-retest reliability observed in the original study. The IRI is a 28-item self-report measure of empathy and so useful for further exploring the EQ's construct validity. It has four subscales, with seven items measured on a 5-point Likert scale ranging from '0 does not describe me well' to '4 describes me very well'. The range of scores for each subscale is 0–35, with 35 representing a high 'empathy' score except on the 'personal distress' scale, which taps self-orientated emotional reactivity.

Results

The test-retest correlation coefficient between EQs administered at time 1 and at time 2 was $r=0.835$ ($n=25$, $p=0.0001$).

The relationship between the IRI and the total EQ score from test time 2, was explored. Moderate correlations were found between the EQ (without the three items that previously correlated with total SDS score) and both the 'empathic concern' ($n=28$, $r=0.423$, $p=0.025$) and the 'perspective-taking' subscale ($n=28$, $r=0.485$, $p=0.009$). The coefficient was $r=-0.027$ for the fantasy items ($p>0.05$) and $r=-0.158$ ($p>0.05$) for the 'personal distress'.

The IRI scores were also correlated with the individual factor scores in order to explore concurrent validity. Factor 2, 'emotional reactivity', showed an association with 'empathic concern' ($n=28$, $r=0.583$, $p=0.001$) and 'perspective taking' ($n=28$, $r=0.442$, $p=0.019$) but not 'personal distress'. Factor 3 ('social skills'), however, displayed a weak but non-significant relationship with perspective taking ($n=28$, $r=0.263$, $p>0.05$). Factor 1, however, did not correlate significantly with any of the IRI subscales.

Study 4

Participants

The DPD group as described in Study 2.

Measures

The EQ and the Dissociative Experiences Scale version II (DES; Bernstein & Putnam, 1986; Carlson & Putnam, 1993) were administered (see Study 2). The DES is the 'gold standard' measure of DPD. It is a 28-item self-report questionnaire with a cut-off score of 30 for severe dissociative disorders (Carlson & Putnam, 1993). Factor analysis suggests three main components: 'depersonalisation/derealisation (DPD/DR)', 'amnesia' for dissociative experiences and 'absorption' and imaginative involvement (Carlson *et al.* 1991). Eight items make up the DES-Taxon which is sensitive to the detection of DPD with a cut-off score of 13 (Simeon *et al.* 1998).

The Beck Anxiety and Depression Inventories (Beck *et al.* 1988*a, b*) were also given to participants due to the co-morbidity between depersonalisation disorder, depression and anxiety (Lambert *et al.* 2001; Baker *et al.* 2003). A score below 11 on either scale is considered within 'normal' range, and a score above 30 is classed as 'severe'.

Table 5. Mean and s.d. EQ scores for the depersonalisation group

	n	Total score on the EQ			
		Mean	s.d.	Min	Max
Male	32	38.9	12.4	15	66
Female	30	46.8	10.1	23	65
Group total	62	42.7	11.9	15	66

Analysis

The mean EQ scores (including all the items) can be found in Table 5. No significant differences were found on total EQ score between the psychologically healthy individuals and those with DPD: for men ($t=1.208$, $df=77$, $p>0.05$) or women ($t=1.496$, $df=90$, $p>0.054$). The difference between men and women with DPD on total EQ scores again reached significance ($t=-2.686$, $df=59$, $p=0.009$).

A 3×2 repeated-measures ANOVA showed a main effect for group [$F(1, 169)=15.11$, $p=0.001$] and a significant effect for factor \times group [$F(2, 338)=12.08$, $p<0.001$]. *T* tests showed the main difference between groups was on 'social skills' with the DPD group rating themselves as less proficient ($t=6.663$, $df=169$, $p=0.001$).

Fifty-three people completed the DES, BAI and BDI. The mean score on the BAI was 21.6 (± 12) and BDI was 20.3 (± 10.5). The mean score on the DES was 23.2 (± 14.2), the DPD/DR subscale 36.6 (± 24), amnesia 6.2 (± 2.5) and absorption 27.6 (± 17). The mean score on the DES taxon was 23.3 (± 15.4). The BAI, BDI and EQ were all entered into a correlational analysis and the co-efficients were found to be close to zero.

Lastly, the relationship between the BDI, BAI and each factor was examined. 'Emotional reactivity' was significantly related to anxiety scores ($n=52$, $r=0.313$, $p=0.024$) and 'social skills' showed a significant negative association with depression scores ($n=45$, $r=-0.346$, $p=0.012$).

GENERAL DISCUSSION

The aim of this study was to examine the reliability, validity and factor structure of the EQ. The mean EQ score was very similar to that found by the original authors indicating the

questionnaires reliability across samples. High test-retest reliability was also shown, as were sex differences mirroring the normative data. However, in Study 1, women scored slightly (but not significantly) higher than the original sample, which may be because a higher proportion of them were drawn from mental health workers. This may prevent firm conclusions regarding sex differences.

The EQ was shown to have concurrent validity as evident from the moderate correlations with the 'empathic concern' and 'perspective-taking' subscales of the IRI (Davis, 1980). The fact that the correlations are only moderate is to be expected, as the total EQ score is an index of global empathy. The weak negative association with 'personal distress' indicates that the two concepts may be inversely related. The lack of association with 'fantasy' items suggests this concept is not empathy *per se* (Baron-Cohen & Wheelwright, in press).

In Study 1, five EQ items correlated significantly with total score on the SDS and three of these were considered of sufficient strength and in the right direction to be left out of the later analysis. In Study 2, one further item was shown to be related to social desirability. The negative correlation with item 37 is somewhat mysterious and may be due to chance factors. That the remaining 35 items showed no association with social desirability supports the scale's construct validity. If the dropped items are to be used in subsequent studies, then it is important to ensure that social desirability is also measured.

The EQ was successfully reduced to a few simple factors which map onto the traditional ideas of empathy, and the final solution accounted for a moderate amount of the total variance. The first factor, cognitive empathy, includes items that measure the appreciation of affective states, i.e. 'I can tell if someone is masking their true emotion', epistemic states, i.e. 'I find it easy to put myself in somebody else's shoes' and desire-based states, i.e. 'I can easily work out what another person might want to talk about'. This is in line with the broader definition of theory of mind as including the attribution of all types of mental state. However, it is also of interest that 'affective state' items had stronger loadings on this factor. This may also explain why no association was found between this factor and the 'perspective-taking'

subscale of the IRI, as the latter is geared more towards epistemic states. Whether or not different types of mental state attribution share the same processes is an issue currently under debate (Stone *et al.* 2003).

The second factor, 'emotional reactivity' reflects the tendency to have an emotional reaction in response to others' mental states, i.e. 'seeing people cry doesn't really upset me'. However, the lack of control for 'personal distress' (Davis, 1980), prevents us from labelling this factor 'emotional/affective' empathy. From these data alone, we cannot be sure that the emotional reactions tapped are other- rather than self-orientated. One way round this is to administer the EQ in conjunction with the 'personal distress' items of the IRI (Davis, 1980). This would give an accurate profile of empathic response. Interestingly, although this factor moderately correlated with 'empathic concern' and 'perspective-taking' on the IRI, it was not associated with 'personal distress' items, suggesting it may be tapping empathy after all.

Sex differences (female superiority) were also found on both cognitive empathy and emotional reactivity but not on the last factor, 'social skills'. This contains items that tap the spontaneous use of such skills and/or a lack of intuitive social understanding, i.e. 'I often find it difficult to judge whether something is rude or polite'. Furthermore, an over-reliance on social rules, i.e. 'I consciously work out the rules of social situations', may be indicative of a lack of spontaneous empathy. Social skills seem to rely on a certain amount of cognitive empathy; hence, the relationship with the perspective-taking subscale of the IRI.

The mean score on the Eyes test matched the normative data. The mean EQ score and 'social skills' had weak but significant correlations with this task. Given that it is an implicit, objective measure of cognitive empathy, this relationship may be important. But the fact that neither association reached significance in the regression analysis needs to be considered. Verbal IQ, as estimated by the NART, was the sole predictor in both of the regression analyses. This raises the possibility that the EQ and Eyes task, in fact, tap different constructs. However, the EQ score approached significance as a predictor in Study 1, and the lack of correlation between

verbal IQ and the EQ suggests they are orthogonal. One possibility is that the role of verbal IQ in the Eyes task was confounded by sample selection. Although possible, this explanation is unlikely, as a dichotomous occupation variable designed to control for this bias was not a predictor of performance. It therefore seems that both verbal IQ and total EQ scores may be related to performance on the Eyes task.

In Study 4, people reporting symptoms of DPD did not suffer a global empathy deficit. They showed the same pattern of EQ scores as the psychologically healthy individuals, including sex differences. The near difference between women in the DPD group and controls is very likely to be an artefact of sample selection (see above). The DPD group did, however, report significantly less ability on items tapping 'social skills'. This association is particularly hard to decipher in the presence of co-morbid depression and anxiety. In line with previous reports, this group tended to score within the mild to moderate range on both the BDI and BAI (Lambert *et al.* 2001). The negative association between 'social skills' and the BDI depression scores may provide one explanation for their lack of social competence. However, no firm conclusions can be drawn without objective measures of 'social skills', which do not rely on potentially biased subjective evaluations, i.e. reports from someone who knows the respondent well. Furthermore, the BAI scores also showed a positive relationship with 'emotional reactivity', emphasizing the need to control for anxiety, as well as depression, when measuring 'affective empathy'. It seems that the effects of both on empathy need to be kept in mind, both in clinical settings and the normal population.

This series of studies confirms that the EQ provides a reliable and valid way of measuring empathy via self-report in both healthy individuals and clinical populations. One limitation is the use of ordinal rather than continuous data in the PCA; however, the factors were easily and meaningfully interpretable. A further limitation is the rather disparate and incompletely characterized samples used including the fact that the sample in Study 1 displayed a high verbal IQ as estimated from the NART. However, the consistency observed across studies suggests that the EQ is robust to such demographic factors.

In terms of future work, it is important to tease out the different kinds of emotional reactivity and distinguish between empathic and other types of emotional responses. Furthermore, the effects of transient states such as anxiety and depression should be taken into account. The EQ would appear to have utility in studying at least two clinical groups: people with Asperger's Syndrome and those with neurotic conditions such as DPD which includes anxiety and depressive symptoms. Further use in clinical research would appear to be worthwhile.

ACKNOWLEDGEMENTS

All staff at the Depersonalisation Research Unit, Maudsley Hospital, Patsy Mott for administrative assistance and everyone who took part in the study. E.L. is supported by an MRC studentship. S.B.-C. was also supported by the MRC. Part of the work was supported by the Pilkington Family Trusts.

NOTE

Supplementary information accompanies this paper on the Journal's website (<http://journals.cambridge.org>).

REFERENCES

- Abdel-Khalek, A., Lester, D. & Barrett, P. (2002). The factorial structure of the Arabic Obsessive-Compulsive Scale in Kuwaiti and American college students. *Personality & Individual Differences* **33**, 3–9.
- Baker, D., Hunter, E., Lawrence, E., Medford, N. C., Patel, M., Senior, C., Sierra, M., Lambert, M. V., Phillips, M. & David, A. S. (2003). Depersonalisation Disorder: clinical features of 204 cases. *British Journal of Psychiatry* **182**, 428–433.
- Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N. & Wheelwright, S. (2003). The Systemising Quotient: an investigation of adults with Asperger's syndrome or high functioning autism, and normal sex differences. *Philosophical Transactions of the Royal Society; Biological Sciences* **358**, 361–374.
- Baron-Cohen, S. & Wheelwright, S. (in press). The Empathy Quotient: an investigation of adults with Asperger's syndrome or high functioning autism, and normal sex differences. *Journal of Autism & Developmental Disorders*.
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y. & Plumb, I. (2001). The 'Reading the mind in the eyes' test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology & Psychiatry & Allied Disciplines* **42**, 241–251.
- Batson, C. D., Fultz, J. & Schoenrade, P. A. (1987). Distress and empathy: two qualitatively distinct vicarious emotions with different motivational consequences. *Journal of Personality* **55**, 19–39.
- Beck, A. T., Epstein, N. & Brown, G. (1988a). An inventory for measuring clinical anxiety: psychometric properties. *Journal of Consulting and Clinical Psychology* **56**, 893–897.

Beck, A. T., Steer, R. A. & Garbin, M. G. (1988b). Psychometric properties of the Beck Depression Inventory: twenty-five years of evaluation. *Clinical Psychology Review* **8**, 77–100.

Bernstein, E. M. & Putnam, F. W. (1986). Development, reliability, and validity of a dissociation scale. *Journal of Nervous & Mental Disease* **174**, 727–735.

Blair, J. (1995). A cognitive developmental approach to morality; investigating the psychopath. *Cognition* **57**, 1–29.

Carlson, E. B. & Putnam, F. W. (1993). An update on the Dissociative Experiences Scale. *Dissociation: Progress in the Dissociative Disorders* **6**, 16–27.

Carlson, E. B., Putnam, F. W., Anderson, G., Clark, P., Torem, M., Coons, P., Bowman, E., Chu, J., Dill, D., Loewenstein, R. J. & Braun, B. G. (1991). *Factor Analysis of the Dissociative Experiences Scale; A Multi-center Study*. Rush Presbyterian: Chicago, IL.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioural Research* **1**, 140–161.

Crowne, D. P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology* **24**, 349–354.

Davis, M. (1980). A multidimensional approach to individual differences in empathy. *Catalog of Selected Documents in Psychology* **10**, 85.

Davis, M. H. (1994). *Empathy: A Social Psychological Approach*. Brown & Benchmark: Dubuque, IA.

Eisenberg, N. & Strayer, J. (eds) (1987). *Empathy and its Development*. Cambridge University Press: Cambridge.

Gorsuch, R. (1974). *Factor Analysis*. W. B. Saunders: Philadelphia.

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology* **33**, 307–316.

Hutcheson, G. & Sofroniou, N. (1999). *The Multivariate Social Scientist*. Sage Publications Ltd: London.

Johnson, J., Smither, R. & Cheek, J. (1983). The structure of empathy. *Journal of Personality and Social Psychology* **45**, 1299–1312.

Kim, J. & Mueller, C. (1978). *Introduction to Factor Analysis*. Sage Publications: Beverley Hills.

Lambert, M. V., Senior, C., Fewtrell, W., Phillips, M. & David, A. S. (2001). Primary and secondary depersonalisation disorder: a psychometric study. *Journal of Affective Disorders* **63**, 249–246.

Mehrabian, A. (2000). Manual for the balanced emotional empathy scale [unpublished]. Available from Albert Mehrabian, 1130 Alta Mesa Road, Monterey, CA 93940, CA, USA.

Mehrabian, A. & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality* **40**, 525–543.

Mehrabian, A., Young, A. L. & Sato, S. (1988). Emotional empathy and associated individual differences. *Current Psychology: Research & Reviews* **7**, 221–240.

Nelson, H. E. (1982). *National Adult Reading Test (NART): Test Manual*. NFER: Wilson, Windsor.

Senior, C., Hunter, E., Lambert, M. V., Medford, N. C., Sierra, M., Phillips, M. L. & David, A. S. (2001). Depersonalisation. *Psychologist* **14**, 128–132.

Shaw, P., Lawrence, E., Baron-Cohen, S. & David, A. S. (2003). The role of the amygdala in social sensitivity. *Annals of New York Academy of Sciences* **985**, 508–510.

Simeon, D., Guralnik, O., Gross, S., Stein, D. J., Schmeidler, J. & Hollander, E. (1998). The detection and measurement of depersonalization disorder. *Journal of Nervous & Mental Disease* **186**, 536–542.

Stone, V. E., Baron-Cohen, S., Calder, A., Keane, J. & Young, A. (2003). Acquired theory of mind impairments in individuals with bilateral amygdala lesions. *Neuropsychologia* **41**, 209–220.

Tabachnick, B. & Fidell, L. (1989). *Using Multivariate Statistics*. Harper Collins: New York.

Appendix 1 – The EQ – (Baron-Cohen and Wheelwright, 2003 in press)

How to fill out the questionnaire

Below are a list of statements. Please read each statement very carefully and rate how strongly you agree or disagree with it by circling your answer. There are no right or wrong answers, or trick questions.

IN ORDER FOR THE SCALE TO BE VALID, YOU MUST ANSWER EVERY QUESTION.

Examples

E1. I would be very upset if I couldn't listen to music every day.	strongly agree	<input checked="" type="radio"/> slightly agree	<input type="radio"/> slightly disagree	<input type="radio"/> strongly disagree
E2. I prefer to speak to my friends on the phone rather than write letters to them.	strongly agree	<input type="radio"/> slightly agree	<input type="radio"/> slightly disagree	<input checked="" type="radio"/> strongly disagree
E3. I have no desire to travel to different parts of the world.	strongly agree	<input checked="" type="radio"/> slightly agree	<input type="radio"/> slightly disagree	<input type="radio"/> strongly disagree
E4. I prefer to read than to dance.	strongly agree	<input type="radio"/> slightly agree	<input type="radio"/> slightly disagree	<input checked="" type="radio"/> strongly disagree
1. I can easily tell if someone else wants to enter a conversation.	strongly agree	<input type="radio"/> slightly agree	<input type="radio"/> slightly disagree	<input type="radio"/> strongly disagree
2. I prefer animals to humans.	strongly agree	<input type="radio"/> slightly agree	<input type="radio"/> slightly disagree	<input type="radio"/> strongly disagree

3. I try to keep up with the current trends and fashions.	strongly agree	slightly agree	slightly disagree	strongly disagree
4. I find it difficult to explain to others things that I understand easily, when they don't understand it first time.	strongly agree	slightly agree	slightly disagree	strongly disagree
5. I dream most nights.	strongly agree	slightly agree	slightly disagree	strongly disagree
6. I really enjoy caring for other people.	strongly agree	slightly agree	slightly disagree	strongly disagree
7. I try to solve my own problems rather than discussing them with others.	strongly agree	slightly agree	slightly disagree	strongly disagree
8. I find it hard to know what to do in a social situation.	strongly agree	slightly agree	slightly disagree	strongly disagree
9. I am at my best first thing in the morning.	strongly agree	slightly agree	slightly disagree	strongly disagree
10. People often tell me that I went too far in driving my point home in a discussion.	strongly agree	slightly agree	slightly disagree	strongly disagree
11. It doesn't bother me too much if I am late meeting a friend.	strongly agree	slightly agree	slightly disagree	strongly disagree
12. Friendships and relationships are just too difficult, so I tend not to bother with them.	strongly agree	slightly agree	slightly disagree	strongly disagree
13. I would never break a law, no matter how minor.	strongly agree	slightly agree	slightly disagree	strongly disagree
14. I often find it difficult to judge if something is rude or polite.	strongly agree	slightly agree	slightly disagree	strongly disagree
15. In a conversation, I tend to focus on my own thoughts rather than on what my listener might be thinking.	strongly agree	slightly agree	slightly disagree	strongly disagree
16. I prefer practical jokes to verbal humour.	strongly agree	slightly agree	slightly disagree	strongly disagree
17. I live life for today rather than the future.	strongly agree	slightly agree	slightly disagree	strongly disagree
18. When I was a child, I enjoyed cutting up worms to see what would happen.	strongly agree	slightly agree	slightly disagree	strongly disagree
19. I can pick up quickly if someone says one thing but means another.	strongly agree	slightly agree	slightly disagree	strongly disagree
20. I tend to have very strong opinions about morality.	strongly agree	slightly agree	slightly disagree	strongly disagree

PLEASE TURN OVER

21. It is hard for me to see why some things upset people so much.	strongly agree	slightly agree	slightly disagree	strongly disagree
22. I find it easy to put myself in somebody else's shoes.	strongly agree	slightly agree	slightly disagree	strongly disagree
23. I think that good manners are the most important thing a parent can teach their child.	strongly agree	slightly agree	slightly disagree	strongly disagree
24. I like to do things on the spur of the moment.	strongly agree	slightly agree	slightly disagree	strongly disagree
25. I am good at predicting how someone will feel.	strongly agree	slightly agree	slightly disagree	strongly disagree
26. I am quick to spot when someone in a group is feeling awkward or uncomfortable.	strongly agree	slightly agree	slightly disagree	strongly disagree
27. If I say something that someone else is offended by, I think that that's their problem, not mine.	strongly agree	slightly agree	slightly disagree	strongly disagree
28. If anyone asked me if I liked their haircut, I would reply truthfully, even if I didn't like it.	strongly agree	slightly agree	slightly disagree	strongly disagree
29. I can't always see why someone should have felt offended by a remark.	strongly agree	slightly agree	slightly disagree	strongly disagree
30. People often tell me that I am very unpredictable.	strongly agree	slightly agree	slightly disagree	strongly disagree
31. I enjoy being the centre of attention at any social gathering.	strongly agree	slightly agree	slightly disagree	strongly disagree
32. Seeing people cry doesn't really upset me.	strongly agree	slightly agree	slightly disagree	strongly disagree
33. I enjoy having discussions about politics.	strongly agree	slightly agree	slightly disagree	strongly disagree
34. I am very blunt, which some people take to be rudeness, even though this is unintentional.	strongly agree	slightly agree	slightly disagree	strongly disagree
35. I don't tend to find social situations confusing.	strongly agree	slightly agree	slightly disagree	strongly disagree
36. Other people tell me I am good at understanding how they are feeling and what they are thinking.	strongly agree	slightly agree	slightly disagree	strongly disagree
37. When I talk to people, I tend to talk about their experiences rather than my own.	strongly agree	slightly agree	slightly disagree	strongly disagree
38. It upsets me to see an animal in pain.	strongly agree	slightly agree	slightly disagree	strongly disagree
39. I am able to make decisions without being influenced by people's feelings.	strongly agree	slightly agree	slightly disagree	strongly disagree

40. I can't relax until I have done everything I had planned to do that day.	strongly agree	slightly agree	slightly disagree	strongly disagree
41. I can easily tell if someone else is interested or bored with what I am saying.	strongly agree	slightly agree	slightly disagree	strongly disagree
42. I get upset if I see people suffering on news programmes.	strongly agree	slightly agree	slightly disagree	strongly disagree
43. Friends usually talk to me about their problems as they say that I am very understanding.	strongly agree	slightly agree	slightly disagree	strongly disagree
44. I can sense if I am intruding, even if the other person doesn't tell me.	strongly agree	slightly agree	slightly disagree	strongly disagree
45. I often start new hobbies but quickly become bored with them and move on to something else.	strongly agree	slightly agree	slightly disagree	strongly disagree
46. People sometimes tell me that I have gone too far with teasing.	strongly agree	slightly agree	slightly disagree	strongly disagree
47. I would be too nervous to go on a big rollercoaster.	strongly agree	slightly agree	slightly disagree	strongly disagree
48. Other people often say that I am insensitive, though I don't always see why.	strongly agree	slightly agree	slightly disagree	strongly disagree
49. If I see a stranger in a group, I think that it is up to them to make an effort to join in.	strongly agree	slightly agree	slightly disagree	strongly disagree
50. I usually stay emotionally detached when watching a film.	strongly agree	slightly agree	slightly disagree	strongly disagree
51. I like to be very organised in day to day life and often make lists of the chores I have to do.	strongly agree	slightly agree	slightly disagree	strongly disagree
52. I can tune into how someone else feels rapidly and intuitively.	strongly agree	slightly agree	slightly disagree	strongly disagree
53. I don't like to take risks.	strongly agree	slightly agree	slightly disagree	strongly disagree
54. I can easily work out what another person might want to talk about.	strongly agree	slightly agree	slightly disagree	strongly disagree
55. I can tell if someone is masking their true emotion.	strongly agree	slightly agree	slightly disagree	strongly disagree
56. Before making a decision I always weigh up the pros and cons.	strongly agree	slightly agree	slightly disagree	strongly disagree
57. I don't consciously work out the rules of social situations.	strongly agree	slightly agree	slightly disagree	strongly disagree

PLEASE TURN OVER

- | | | | | |
|---|-------------------|-------------------|----------------------|----------------------|
| 58. I am good at predicting what someone will do. | strongly
agree | slightly
agree | slightly
disagree | strongly
disagree |
| 59. I tend to get emotionally involved with a friend's problems. | strongly
agree | slightly
agree | slightly
disagree | strongly
disagree |
| 60. I can usually appreciate the other person's viewpoint, even if I don't agree with it. | strongly
agree | slightly
agree | slightly
disagree | strongly
disagree |

Thank you for filling this questionnaire in.

© SBC/SJW Feb 1998