Development and validation of the Basic Empathy Scale

Darrick Jolliffe*, David P. Farrington

Institute of Criminology, University of Cambridge, Sidgwick Site, Cambridge, England CB3 9DT, UK

Abstract

In developing the Basic Empathy Scale (BES), 40 items measuring affective and cognitive empathy were administered to 363 adolescents in Year 10 (aged about 15). Factor analysis reduced this to a 20-item scale that was administered 1 year later to 357 different adolescents in Year 10 in the same schools. Confirmatory factor analysis verified the two-factor solution. Females scored higher than males on both affective and cognitive empathy. Empathy was positively correlated with intelligence (for females only), extraversion (cognitive empathy only), neuroticism (affective empathy only), agreeableness, conscientiousness (for males only), and openness. Empathy was positively related to parental supervision and socioeconomic status. Adolescents who would help victims of bullying had high empathy.

Introduction

Usually defined as an affective trait (e.g. the capacity to experience the emotions of another; Bryant, 1982) and/or a cognitive ability (e.g. the capacity to comprehend the emotions of another; Hogan, 1969), the concept of empathy is considered important in many different areas of psychology (e.g. Strayer, 1987). In particular, empathy and the acquisition of empathy are considered essential components of adequate moral development. In line with this assertion, researchers have demonstrated empirical relations between many forms of prosocial behaviour and empathy (e.g. Batson, Fultz, & Schoenrade, 1987).

*Corresponding author. Tel.: +44 1223 335360; fax: +44 1223 335356.
E-mail address: dj211@cam.ac.uk (D. Jolliffe).
Following from the relationship between empathy and prosocial behaviour, it has been postulated that a lack of empathy may be associated with aggressive and antisocial behaviour. This is because individuals who share and/or comprehend another’s negative emotional reaction (e.g. distress), which occurs as a result of their own aggressive or antisocial behaviour, may be inhibited and less inclined to continue with this behaviour or act in an antisocial or aggressive manner in the future (Feshbach, 1975).

In order to test the relationship between empathy and aggressive/antisocial behaviour, Miller and Eisenberg (1988) conducted a meta-analysis of 43 studies. Across these studies, empathy was assessed in a number of ways including picture/story presentations, questionnaires, facial and gesture responses and behavioural responses to experimental induction. Empathy was operationalized in purely affective terms (i.e. emotional responses evoked by the affective state or situation of the other person: see their p. 325). Aggressive/externalizing behaviour was also defined broadly to include self-report measures of aggression, peer/teacher ratings of aggression and administration of ‘shock’ to an experimental confederate.

Negative correlation coefficients in the low to moderate range (−0.06 to −0.46) were reported between empathy and aggression/externalizing behaviour. However, a significant relationship \( r = -0.18 \) was found when empathy was assessed using questionnaires. This was attributed to the age of the subjects, as questionnaires were usually given to adolescents and adults, who in the past have shown more consistent relationships between empathy and behaviour (e.g. Eisenberg & Miller, 1987).

Although the above results provide limited support for a negative relationship between empathy and antisocial behaviour, a more recent systematic review and meta-analysis of empathy and criminal offending (Jolliffe & Farrington, 2004) calls into question the strength this relationship. This review identified 35 studies of empathy and offending. Empathy was operationalized as responses to questionnaire measures of empathy (both affective and cognitive) and offending was operationalized as delinquent or criminal behaviour (mostly based on official records).

A moderate mean effect size \( (d) = -0.27 \) (approximately equal to an \( r \) of \(-0.14\)) was found, suggesting a negative relationship between empathy and offending. This relationship was stronger for cognitive empathy than for affective empathy, and stronger for younger people compared to older people. However, the most striking finding was that the relationship between empathy and offending was reduced considerably after controlling for intelligence and disappeared completely after controlling for socio-economic status. This suggests that either (1) empathy is not causally linked to offending, because low socio-economic status causes both low empathy and offending, or (2) low socio-economic status causes low empathy which in turn causes offending.

There were three main scales identified and used in the Jolliffe and Farrington (2004) systematic review and meta-analysis. These were the Hogan Empathy Scale (HES) (Hogan, 1969), the Questionnaire Measure of Emotional Empathy (QMEE)\(^1\) (Mehrabian & Epstein, 1972) and the Interpersonal Reactivity Index (Davis, 1980). The HES is considered to measure cognitive empathy, the QMEE is considered to measure emotional empathy and the IRI is considered to measure both cognitive and emotional empathy.

\(^1\)Both the Impulsiveness-Venturesomeness-Empathy Scale (Eysenck, Easting, & Pearson, 1984) and the Index for Empathy in Children and Adolescents (Bryant, 1982) were derived from the QMEE.
Interestingly, the meta-analysis demonstrated that these scales showed considerable variation in their ability to detect empathy differences between offenders and non-offenders. Upon inspection it was determined that the scale which initially appeared the ‘best’ at differentiating offenders and non-offenders (HES), did so, not because of a superior ability to detect empathy, but because of a potential artefact: the items were not face valid items which could plausibly measure empathy, but were chosen according to their ability to discriminate between previously categorized high and low empathy groups. In fact, it is unlikely that this scale measures empathy. Before the causal relationships between empathy and offending can be properly tested in prospective longitudinal studies, a reliable and valid measure of empathy is needed.

The remaining scales available to measure empathy, the QMEE and the IRI, appear to have shortcomings which might limit their usefulness. First, both appear to equate sympathy with empathy. Second, neither of these scales yield a measure of cognitive empathy.2 Third, while these scales are commonly used to assess empathy levels in offenders and similar populations, both were developed and validated using university undergraduates. These weaknesses and the implications of these weaknesses will be discussed in turn.

An inspection of the definitions of empathy used for the development of the QMEE and IRI, and the items on these scales, indicate that they may, in fact, be failing to measure empathy adequately. The definition of empathy used for the Mehrabian and Epstein (1972) QMEE was “…a vicarious emotional response to the perceived emotional experiences of others” (p. 523), which appears to fit well with the current conceptualization of affective empathy. However, an investigation of the items used to measure empathy indicates that this scale is not measuring emotional empathy but rather “…related aspects of emotional empathy” (p. 525). Items such as: “I get very angry when I see someone being ill-treated” (agree) and “I become more irritated than sympathetic when I see someone in tears” (disagree), are clearly measuring sympathetic or concerned feelings for others, rather than shared emotion. Similarly, the EC scale of the Davis (1980) IRI is designed to assess “other-oriented feelings of sympathy and concern for unfortunate others” (p. 114), with items such as “Other people’s misfortunes do not usually disturb me a great deal” (disagree) and “I am often touched by things that I see happen” (agree).

It appears that these scales may equate sympathy with empathy. However, these are distinct and separable constructs (Feshbach, 1975; Eisenberg & Strayer, 1987). Like empathy, sympathy has been defined in many ways, but currently there appears to be a general consensus that sympathy involves the appraisal of how one feels about the emotions of another (e.g. Eisenberg & Strayer, 1987). Defined in this way sympathy is similar to affective empathy in that both involve an affective reaction to the perceived emotions of another. However, in the case of affective empathy this reaction is the same emotion as the target person (emotion congruence), whereas in sympathy this affective reaction may not necessarily be the same emotion. Sympathy might also be considered similar to cognitive empathy as both would be expected to involve an understanding of the emotional situation of another person. Sympathy however, would involve an additional appraisal regarding this emotional understanding. For example, one may share or understand another’s sadness (affective/cognitive empathy) and then feel concerned about that person’s welfare (sympathy). Alternatively, one may share or understand another’s

---

2The perspective taking scale of the Interpersonal Reactivity Index is often treated as a measure of cognitive empathy, but it will be suggested that this scale does not sufficiently assess the understanding of another person’s emotions.
sadness (affective/cognitive empathy), but feel that that person deserves their situation (no sympathy).

In many studies it is not clear if it is sympathy or empathy that is being referred to when ‘empathy’ is related to antisocial behaviour (e.g. Strayer, 1987). Both sympathy and empathy would be expected to decrease the likelihood of antisocial or offending behaviour. While it might be expected that sympathy (involving the additional appraisal) would be more closely related to antisocial behaviour than empathy (e.g. Hanson, 2003), it is difficult to imagine a situation where sympathy could arise without some degree or form of empathy.

Obviously, it is important to tease these three concepts apart. For example, it may be that offenders have sufficient levels of cognitive and affective empathy, but low levels of sympathy, which increases their likelihood of antisocial behaviour. If true, treatment designed to increase empathy would not have the desired effect. Treatment to increase sympathy might be more effective. Alternatively, offenders may have low cognitive or affective empathy and low sympathy. In this case empathy treatment would be necessary to reduce antisocial behaviour, but would not be sufficient to reduce offending until sympathy was also increased.

Another difficulty with the present empathy questionnaires is that they do not measure cognitive empathy. As previously stated, the HES should no longer be considered a useful measure of cognitive empathy, and the PT scale of the IRI measures perspective taking, which is similar to, but clearly not the same as, cognitive empathy. Items such as: “When I am upset at someone, I usually try to ‘put myself in his shoes’ for a while” (agree) and “I sometimes find it difficult to see things from the ‘other guy’s’ point of view” (agree), illustrate that this scale measures the broad ability to take another’s perspective, rather than the more specific ability to understand the emotions of another.

Having a valid measure of cognitive empathy is essential for understanding the relationship between empathy and offending as both affective and cognitive empathy would be expected to provide a unique contribution to this relationship. This can be illustrated by imagining a situation whereby a person possesses a deficiency in only one component of empathy. For example, it has been suggested that the superficial charm of psychopaths results from having sufficient cognitive empathy, but their willingness to act without regard to the other’s feelings demonstrates their lack of affective empathy (Strayer, 1987; Tangney & Stuewig, 2004). It would be useful to have a device that would allow these hypotheses to be tested.

Another difficulty with the QMEE and the IRI is that they were both validated using similar populations (e.g. university students). This poses a problem when these scales are used to assess empathy in more heterogeneous populations as the higher intelligence and introspective abilities of university students may influence what is actually being measured. Furthermore, university students would be of an age, social status and family background which would make them less likely than others to be involved in offending ever or currently. Therefore, the aspects of low empathy that might be important for offending may be missed.

The main aim of this paper is to describe the development of a new measure of affective and cognitive empathy that will attempt to overcome the shortcomings of the existing questionnaires. The new measure called the Basic Empathy Scale (BES) was based specifically on the definition of empathy put forth by Cohen and Strayer (1996) “as the understanding and sharing in another’s emotional state or context” (p. 523). This orientation was adopted because it allowed for a focus
on both affect congruence (affective empathy) and the understanding of another’s emotions (cognitive empathy).

**Generation of items**

Items were generated based on the above conceptualization of affective and cognitive empathy. For example, the item: “I usually feel calm when other people are scared” (disagree) was designed to assess affective empathy by the degree to which the subject shares fear with others. Emphasis was placed on identifying and designing affective empathy items which assessed emotion congruence. In this way the common overlap between affective empathy and sympathy can be reduced substantially; responses to these items should reflect only empathy.

Similarly, the item “It is hard for me to understand when my friends are sad” (disagree) was designed to assess cognitive empathy by the degree to which subjects understand their friend’s sadness. The emphasis of these items is on the understanding of another’s emotions rather than the understanding of another’s perspective.

Emotive words and phrases used to imply empathy (e.g. feeling sorry for another) were avoided in item selection. These words and phrases are value-laden and increase the risk of presentation bias as they can be easily seen through by participants. Kline (1993) suggests that, in order to reduce this potential bias of self-presentation, items should be designed that reduce the insight that subjects might have into the trait under investigation. In this way more accurate responding may be facilitated. Arguably, items that match the current definition of empathy are less likely than those on previous scales to provoke a self-presentation bias. This is because experiencing emotions which correspond to those of another, and to a lesser extent understanding the feelings of another, do not have the same social expectations as sympathy (Feshbach, 1975; Strayer, 1987). There is a socially desirable expectation to feel some unfavourable emotion (e.g. sadness) for a person in an unfortunate situation, and thus subjects who may not feel this way may feel they need to modify their response to fit with this. Although these new items may not completely eliminate self-presentation, what constitutes the socially desirable answer may be less clear on items of the BES scale.

Items for the BES were also based on four of the five ‘basic emotions’ (fear, sadness, anger, happiness). It has been proposed that all emotions derive from these basic emotions (Power & Dalgliesh, 1997). It was hoped that by maintaining a focus on basic emotions the experience of cognitive and affective empathy might be superior to that of non-specific emotions such as being nervous, anxious or upset, commonly used in other scales. It has been suggested that these non-specific emotions may be more ambiguous or require additional and possibly idiosyncratic appraisals (Power & Dalgliesh, 1997). Eight general empathy items were also included (four affective and four cognitive). Examples of these items are “I get caught up in other people’s feelings easily” (affective) and “I can often understand how people are feeling even before they tell me” (cognitive).

Based on the above criteria a total of 40 items were selected. These items were counterbalanced to overcome possible acquiescence response bias with 20 requiring a positive response and 20 requiring a negative response. Each item asked the participant to respond on a Likert scale from 1 representing ‘strongly disagree’ to 5 representing ‘strongly agree’, depending on the degree to which the item described them.
Method

The sample

The 40 item BES scale was initially administered to a sample of 363 adolescents (194 males, 169 females) in Year 10 (mean age 14.8, S.D. = 0.48) in three schools in Hertfordshire, England. This age group was specifically chosen because the Jolliffe and Farrington (2004) meta-analysis demonstrated that young people have stronger relationships between low empathy and offending than adults. This fact, along with studies that have shown that mid-adolescence is the period of highest prevalence and frequency of offending (e.g. Loeber et al., 2003), suggest that empathy deficits may play a crucial role in this time period. Also, by developing a scale on a group of youths with diverse social and intellectual backgrounds this scale may more accurately reflect empathy levels among all young people. The use of adolescents might also make it more likely that the scale will be of a reading and comprehension level suitable for use with adult antisocial populations.

Procedure

The 363 participants rated themselves on the 40 item BES using the five-point response scale. Also, in order to assess the construct validity of the BES, participants were requested to complete a number of additional measures. These were:

Sympathy and perspective taking: The Interpersonal Reactivity Index (Davis, 1980). While the original IRI is composed of four, seven-item scales together designed to assess “reactions of one individual to the observed experiences of another” (Davis, 1983, p. 113), it was decided that only the Perspective Taking (PT) and Empathic Concern (EC) scales would be used. The PT scale assesses the tendency to spontaneously adopt the psychological point of view of others. The EC scale assesses ‘other oriented’ feelings of sympathy and concern for unfortunate others. In a sample of 460 university students (225 males, 235 females), Davis (1983) found that the EC scale correlated with the QMEE $r = .63$ for males and $r = .56$ for females. Similarly, the PT scale correlated with the QMEE $r = .22$ for males and $r = .17$ for females.

The Cronbach alphas for this scale for the sample was .67 for perspective taking and .70 for sympathy (alpha males = .65 and .67; alpha females = .67 and .65).

Alexithymia: The Toronto Alexithymia Scale —20 (TAS; Bagby, Parker, & Taylor, 1994) was used to assess alexithymia in this sample. Alexithymia refers to a condition whereby one has a diminished ability to identify and/or communicate feeling. Example items would be ‘I often don’t know when I am angry’ (agree) and ‘I am often confused about what emotion I am feeling’ (agree).

In their study of 100 university undergraduates (87 of which were female) Davies, Stankov, and Roberts (1998) found that alexithymia was negatively but weakly related to both the QMEE and the HES. This makes conceptual sense as those with a poor ability to identify their own feelings would be expected to possess a diminished ability to comprehend, and possibly also to experience

---

3That is, there is a greater difference in empathy between young people who offend and young people who do not than the difference in empathy between adults who offend and those who do not.
another’s emotions. Similarly, while difficulty in communicating feelings might not influence the experience of cognitive and affective empathy, this might have a negative impact on one’s responses to measures of empathy.

The Cronbach alpha for this sample was .71 (alpha males = .71, alpha females = .72).

**Intelligence (verbal fluency):** The Thurstone Verbal Fluency Task was included as a measure of intelligence. This task requires subjects to name as many words in a given category as possible in 1 min. This measure was one of two verbal components included in Thurstone’s (1938) conception of intelligence, and verbal fluency continues to be an important component of present day intelligence theories and tests (e.g. Carroll, 1993). This test has been shown to differentiate patients with frontal lobe damage (e.g. Milner, 1964). Furthermore, both PET (Frith, Friston, Liddle, & Fracknowiak, 1991) and functional magnetic resonance imaging (fMRI) scans (Phelps, Hyder, Blamire, & Shulman, 1997) indicate, that the left frontal cortex is activated during verbal fluency tests.

Few studies have investigated the possibility that intelligence may influence cognitive and/or affective empathy. However, the Jolliffe and Farrington (2004) meta-analysis found that offender/non-offender comparison studies which controlled for intelligence had significantly lower mean effect sizes than those that did not. Davis (1983) compared the IRI to scores on the Wechsler Intelligence test for 60 males and 54 females. Correlations between intelligence and empathy scales ranged from −.22 for females to +.15 for males. None of these correlations was significant, likely because of the small numbers. Similarly, Davies et al. (1998) found correlations between six short intelligence measures and affective empathy (QMEE) ranging from −.05 to +.20 (only one significant) and −.04 to +.27 (only two significant) when compared to cognitive empathy (HES).

The mean score of this test was 14.4 (s.d. = 4.1) for the entire sample with scores ranging from 1 to 28. In line with previous research regarding the greater verbal ability of females (e.g. Moffitt, Caspi, Rutter, & Silva, 2001), females were found to score significantly higher than males on this measure (15.3 compared to 13.5, t = 6.0, p < .0001). However, there was no significant change in the mean of this measure over the two waves of data collection (14.4 (s.d. = 4.0) compared to 14.4 (s.d. = 4.1) t = −0.2, n.s.) adding further support to the reliability of this measure.

**Impulsivity:** The 12-item urgency scale of the UPPS Impulsivity Scale (Whiteside & Lynam, 2001) was used to assess impulsivity in this sample. High scorers on the Urgency scale ‘are likely to engage in impulsive behaviors in order to alleviate negative emotions despite the long-term harmful consequences of these actions’ (Whiteside & Lynam, 2001, p. 685). A typical item is: “It is hard for me to resist acting on my feelings” (agree).

Impulsivity is often considered one of the most important features of antisocial behaviour (e.g. White et al., 1994). For example, numerous prospective longitudinal studies have found that measures of impulsivity (or variants of this construct) significantly predicted future offending (e.g. Farrington, 2003; Loeber et al., 2003; Moffitt et al., 2001). Studies of empathy and offending have rarely considered the moderating influence that impulsivity might exert on this relationship (for an exception see Luengo, Otero, Carrillo-De-La-Pena, & Miron, 1995). That is, high impulsivity may over-ride high empathy in the commission of an offence, or high impulsivity and low empathy may both be related to offending because they both reflect a third variable such as poor executive brain function.

This scale produced a Cronbach alpha of .88 for the entire sample (alpha males = .88, alpha females = .87).
Personality: The big-five inventory (BFI; John & Srivastava, 1999) was used to assess personality in this sample. This 44 item scale was designed to assess the five domains of personality (extraversion (E), neuroticism (N), agreeableness (A), conscientiousness (C) and openness (O)) by asking respondents to indicate how much each of the statements describes them. E refers to someone who seeks excitement and is sociable; A refers to someone who is altruistic and sympathetic; C refers to someone who is thorough and organized; O refers to someone who is imaginative and excitabale; N refers to someone who is self-conscious and anxious (John & Srivastava, 1999). Typical items included: ‘I see myself as someone who tends to be disorganised’ (conscientiousness; disagree) and ‘I see myself as someone who worries a lot’ (neuroticism, agree).

Measures of empathy have rarely been compared to omnibus measures of personality such as the five-factor model. This is likely because empathy is considered a component of many personality scales. For example, empathy is part of the altruism component of agreeableness, which is itself one of the factors of the five factor model of personality (John & Srivastava, 1999). More commonly empathy measures have been compared to other scales designed to measure other distinct components of personality. For example, a comparison of dispositional shyness (a component of neuroticism) to the IRI found correlations of $r = -.20$ for females and $r = -.13$ for males for the PT scale and $r = .07$ for females and $r = .15$ for males for the EC scale (Davis, 1983). Similarly, when the IRI was compared to a measure of extraversion, correlations were $r = .05$ for females and $r = .12$ for males for the PT scale and $r = .07$ for females and $r = -.09$ for males for the EC scale.

The Cronbach alpha coefficients for the entire sample were found to be: $E = .76$, $A = .68$, $C = .74$, $O = .71$, $N = .74$ (Males alpha $E = .76$, $A = .65$, $C = .73$, $O = .69$, $N = .68$, Females alpha $E = .76$, $A = .69$, $C = .75$, $O = .73$, $N = .75$).

Socioeconomic status (SES): SES was assessed by asking the children if their parents were employed and also their type and place of employment. A value from 1 (e.g. teachers and other professionals) to 8 (long-term unemployed) was assigned according to the system developed by the Office of National Statistics for the 2001 Census (Walker et al., 2002). In order to generate a single measure of SES for each child, the higher value of either maternal or paternal SES was used. This also compensated for the difficulty presented by those children with only a mother or a father at home.

Little research has been devoted to the relationship between SES and empathy. The Jolliffe and Farrington (2004) meta-analysis demonstrated that, in offender/non-offender comparison studies, those studies that controlled for SES did not find a significant relationship between empathy and offending. This indicates that there may be a relationship between empathy and SES that is worth investigating.

The mean score for this measure was 3.6 (s.d. = 1.8) for the entire sample. There were no significant SES differences between the three schools ($F = 2.8$, n.s.), the two waves of data collection ($F = 0.90$, n.s.) or between males and females ($F = 3.1$, n.s.).

Parental supervision: Four questions were used to assess the typical parenting that the child received. For example, ‘When you are out in the evenings by yourself or with friends do your parents (or step-parents) tell you what time to return home?’ These four items were scored on a five point Likert scale and summed to produce a measure of poor parental supervision.

Much research has been devoted to the relationship of parenting to the development of empathy in children (e.g. Chase-Lansdale, Wakschlag, & Brooks-Gunn, 1995). It has been
suggested that sufficient supervision of the child by the parent is essential for normal emotional
development to occur (e.g. Bowlby, 1982). It would therefore be expected that low parental
supervision would be related to low empathy.

Cronbach’s alpha for this sample was .72 (alpha males = .71, alpha females = .74). In line with
expectation (e.g. Moffitt et al., 2001), females reported significantly better supervision than males
(mean females = 8.1 (S.D. = 2.3) compared to mean males = 8.8 (S.D. = 2.3), t = 3.7, p < .0001).

Social desirability scale: Questionnaire measures of empathy have been criticized for being open
to presentation bias and, therefore, a more accurate reflection of how empathic the person wishes to
be seen as rather than how empathic they actually are (Eisenberg & Fabes, 1990). In order to assess
this bias, six items from the Lie scale of the Eysenck personality questionnaire (Eysenck & Eysenck,
1991) were included as a measure of social desirability. An example item is “I sometimes boast a little”
(disagree). Cronbach’s alpha for this sample was .62 (alpha males = .61, alpha females = .63).

Behavioural response to witnessing bullying: Each respondent was asked about their typical
response to seeing someone bullied in school.

As previously mentioned high empathy is proposed to increase the likelihood of prosocial
behaviour. It is postulated that this is because those who can share or understand another’s
distress will be motivated to assist that person to reduce their own experience of distress. This is
supported by numerous experimental studies which have shown that subjects manipulated to feel
more empathy are more likely than others to help an unfortunate person (Batson et al., 1987).

Data reduction

The goal of the data reduction was to develop a concise and coherent scale which measured
cognitive and affective empathy. For this reason an exploratory factor analysis (principal
component analysis and varimax rotation) of the responses of the 363 participants to the 40 item
BES scale was undertaken. As the items were chosen to assess two specific factors of empathy, a
two factor solution was forced on the data. It is more common for an exploratory factor analysis
to examine all the factors evident when the eigenvalues are >1. However, when this was
undertaken a large number of non-sensical factors were extracted. An examination of the scree
plot (e.g. Cattell, 1966), however, indicated that the instrument’s structure was more
appropriately described as having two factors, thus supporting this decision.

Factor 1 (affective empathy) accounted for 19.5% of the variance and Factor 2 (cognitive
empathy) accounted for 7.6% of the variance. Of the 20 cognitive items administered 9 were
selected for the final cognitive scale (alpha = .79), and of the 20 affective items administered, 11
were selected for the final affective scale (alpha = .85). Items were not included in the final scale if
they did not load above 0.40 on either factor. As cognitive and affective empathy were expected to
be correlated, items were not deleted if they loaded on both factors, as long as they loaded more
heavily in the theoretically expected direction.

A second wave of data collection was undertaken exactly one year later with 357 different
respondents (182 males, 175 females) from the same schools and in the same year group. All of the
same measures were used, except that the shortened 20-item BES (based on factor analysis) was
used (nine cognitive items and 11 affective items). This means that while the initial factor analysis

---

4The social desirability scale was given only in the second wave of data collection.
of the BES was only based on 363 respondents, the validation of the scale was conducted on the foil sample of 720 respondents.

A confirmatory factor analysis was undertaken on the 20-item BES based on the entire sample of 720 students. Fig. 1 shows that the loadings for the cognitive items ranged from 0.43 to 0.62 and that the loading for the affective items ranged from 0.41 to 0.71. Following the recommendations of Cole (1987), Cuttance and Ecob (1987), and Marsh, Balla, and McDonald (1988), the goodness-of-fit of the final empathy scale was evaluated using multiple criteria: the goodness-of-fit index (GFI), the adjusted goodness-of-fit index and the root mean square residual (RMS, Joreskog & Sorbom, 1986). Multiple criteria were used because each index has different strengths and weaknesses in assessing goodness-of-fit between a particular model and the observed data. Based on the recommendations of Anderson and Gerbing (1984), Cole (1987), Cuttance and Ecob (1987) and Marsh et al. (1988), the following criteria were used to assess goodness-of-fit: GFI > 0.85, AGFI > 0.80 and RMS < .10.

All three measures of goodness-of-fit indices suggest that the separation of the cognitive and affective factors was a good fit to the data. The GFI was found to be 0.89, the AGFI was found to be 0.86 and the RMS was 0.06. The suitability of a unidimensional model was also tested. The GFI was found to be 0.82, the AGFI was found to be 0.78 and the RMS was 0.08. The first two of these figures are lower than recommended and lower than those produced by a two-factor solution, suggesting the two-factor solution is superior.

In order to further test the structural validity of the BES confirmatory factor analysis was undertaken separately for males and females. When the two-factor solution was investigated the GFI for males was 0.88 and 0.86 for females, the AGFI for males was 0.85 and 0.83 for females, and the RMS for males was 0.07 and 0.06 for females. The corresponding results for the one-factor solution were: GFI = 0.79 for males, 0.81 for females; AGFI = 0.74 for males, 0.76 for females; RMS = 0.09 for males and 0.08 for females. These results provide further support for the two factor solution.

**Testing the construct validity of the BES**

Theory and empirical evidence were used to develop a number of predictions regarding the BES and the other measures collected. The following are a list of the predictions and the basis on which these predictions were made.

*Females should score significantly higher than males on the BES*

Research has consistently found that females score much higher on measures of empathy than males (e.g. Davis, 1983; Lennon & Eisenberg, 1987), especially for questionnaire measures of empathy. It is unclear whether this difference represents a ‘true’ difference (i.e. the result of females being socialized to be more responsive than males to the feelings of others) or is a result of biased responding in self-reports (i.e. because females are expected to be more responsive to the feelings of others, they respond to questionnaires in concordance with this sex-role stereotype). It might be expected that the differences in empathy between males and females might be greater for affective than for cognitive empathy.
Fig. 1. Two-factor confirmatory factor analysis model for the entire sample.
There should be a significant positive correlation between affective and cognitive empathy as measured by the BES

Although affective and cognitive empathy have been considered to be separable processes for some time (e.g. Hills, 2001), both involve reactions to the emotions of another. Most researchers (e.g. Hoffman, 1987; Marshall, Hudson, Jones, & Fernandez, 1995; Strayer, 1987) believe that affective empathy leads to cognitive empathy. That is, the experience of another’s emotions (affective empathy) through emotional contagion is the basis of empathy and that this experience results in the cognitive understanding of these emotions (cognitive empathy). Regardless of the possible causal order, a degree of overlap between cognitive and affective empathy would be expected.

Significant positive correlations between the scales of the BES and sympathy and perspective taking as measured by the IRI would be expected

As both the IRI and the BES have been designed to measure emotional responses to others and understanding the perspectives/emotions of others it would be expected that these measures would correlate significantly. A previous comparison of the EC and PT scales of the IRI to other measures of empathy (the HES and the QMEE) identified correlations ranging from $r = .11$ to $r = .63$ (all $p < .05$; Davis, 1983).

A significant negative correlation between the BES and the measure of alexithymia would be expected

Previous research (e.g. Davies et al., 1998) has suggested that empathy and alexithymia are negatively related. Also, as previously mentioned, this makes conceptual sense as those with a poor ability to identify their own feelings might be expected to possess a diminished ability to comprehend and possibly also to experience another’s emotions. That is, high alexithymia might block affective or cognitive empathy.

A significant positive correlation between the BES and the measure of verbal fluency would be expected

This prediction was based on previous research which has found positive correlations between various measures of intelligence and empathy (Davies et al., 1998; Davis, 1983). While the correlations found in these studies were not strong it was still thought that those with a greater ability to express themselves verbally might have greater access to their emotions and therefore display both greater cognitive and affective empathy.

A significant negative correlation between the BES and impulsivity would be expected

It was not possible to identify any studies which compared measures of empathy to measures of impulsivity on which to based this prediction. However, the above prediction was based on the assumption that those who act without considering the consequences of their actions (high
impulsivity) will also demonstrate a decreased ability to understand and/or experience the
emotions of others. That is, those with fewer cognitive facilities (e.g. executive function deficits)
might have both high impulsivity and low empathy. Also, research has suggested that high
impulsivity and low empathy are both related to offending (Luengo et al., 1997), suggesting a
possible inverse relationship.

**Significant positive correlations would be expected between the BES and the personality clusters agreeableness, openness and conscientiousness, and negative correlations with neuroticism and extraversion**

It was not possible to identify any studies which compared measures of empathy to the five-
factor model of personality. These predictions were based on comparisons of empathy to other
omnibus measures of personality (e.g. Davis, 1983) and the definitions of the five factors of
personality. For example, agreeableness refers to one who altruistic and sympathetic, both of
which are proposed to be related to empathy (e.g. Batson et al., 1987). Similarly, neuroticism
refers to one who is self-conscious and anxious. Some research has suggested that anxiety might
interfere with empathy (e.g. Lindsey, Carlozzi, & Eells, 2001).

**Significant negative correlations between the BES and poor parental supervision and low SES would be expected**

This prediction was based on research which has suggested that those who receive high-quality
parenting have greater empathy than those who do not (e.g. Laible, Carlo, & Roesch, 2004).
Therefore, poor parenting should be negatively related to empathy. While no evidence exists with
which to connect empathy and SES, longitudinal research has generally found low SES and poor
parenting to be positively related (Farrington, 2003). Low SES may result in stressful
circumstances, which might result in poor parenting, in turn resulting in low empathy.

**The BES will not be significantly related to social desirability**

A valid measure of empathy should measure a subject’s empathy rather than how empathic a
person wishes to be perceived. Previous self-report measures of empathy have been criticized for
being open to presentation bias (e.g. Eisenberg & Fabes, 1990), but few empirical investigations
have addressed this topic. Those that have addressed this have found no relationship between the
measures of empathy and social desirability (e.g. Davis, 1983; Mehrabian & Epstein, 1972).

**Those who provide assistance to children being bullied in school should score higher on the BES than those who do not**

As previously mentioned high empathy is proposed to increase the likelihood of prosocial
behaviour. This is because those who can share or understand another’s distress should be
motivated to assist that person to reduce their own experience of distress. This is supported by
numerous experimental studies which have shown that subjects manipulated to feel more empathy
are more likely than others to help an unfortunate person (Batson et al., 1987).
Results

Where applicable, effect sizes in addition to statistical significance will be used to assist in the interpretation of the results. It is important not to rely on statistical significance as this can indicate a large effect in a small sample or a small effect in a large sample. Therefore, a significant result does not necessarily indicate a strong effect, and alternatively a non-significant result does not necessarily indicate a weak effect. A number of methods of interpreting the magnitude of effect sizes have been proposed. A widely used convention is that proposed by Cohen (1988). An effect size less than about .20 is considered small, while an effect size around .50 is considered medium and an effect size greater than about .80 is considered large. These criteria are probably too stringent.

A more meaningful way of interpreting the effect size can be provided by converting the results to the differences in proportions between two groups. First, the effect size \( (d) \) is converted to a phi correlation \( (r) \) (Lipsey & Wilson, 2001, p. 199). This results in an \( r \) value of approximately half that of \( d \), and this value of \( r \) or phi is, in turn, approximately equal to the difference in proportions between the two groups (Farrington & Loeber, 1989). For example, if 50% of offenders had low empathy, compared with 25% of non-offenders, phi would be about .25 and \( d \) would be considered medium (.50), but this seems a substantial difference.

Gender differences in empathy

For both the cognitive and affective scales of the BES females scored significantly higher than males. For the cognitive scale, males had a mean of 32.2 (S.D. = 5.1), and females had a mean of 35.0 (S.D. = 3.9) \( (t = 8.3, p < .0001, d = 0.63) \). For the affective empathy scale males had a mean of 32.1 (S.D. = 6.5), while females had a mean of 40.3 (S.D. = 5.8) \( (t = 17.6, p < .0001, d = 1.33) \). There was also a significant difference in the total score of the BES between males and females (mean = 64.3 (S.D. = 9.8) for males compared to 75.3 (S.D. = 8.3) for females, \( t = 16.1, p < .0001, d = 1.22 \)). Therefore, in line with expectation females clearly scored much higher than males on the all scales of the BES and the magnitude of the difference between males and females was greater for the affective than the cognitive scale.

Relationship between cognitive and affective empathy

The correlation between the affective and cognitive scales of the BES was found to be \( r = .41 \) \( (p < .0001) \) for males and \( r = .43 \) \( (p < .0001) \) for females. This suggested a significant amount of overlap between the cognitive and affective components of empathy for both males and females, but also highlights that a degree of differentiation existed. Further examination suggested that the moderate correlation between these scales was the result of the loading of three affective items on the cognitive scale. Items E1, E2 and E18 were correlated with the affective scale at \( r = .46, r = .43 \) and \( r = .45 \), respectively.5

5Maintaining these three items as affective items was supported by the finding of higher correlations when compared to the affective scale with that item removed. Correlations for items E1, E2 and E18 to the affective scale were \( r = .53, r = .64 \) and \( r = .61 \), respectively.
Table 1 shows the Pearson Correlations between the cognitive, affective and total scales of the BES and the theoretically related measures for males and females separately.

The similarity between the total of the BES and the total of the IRI for males and females was reflected in their relatively high inter-correlations \((r = .53 \text{ males}, \ r = .43 \text{ females})\). This clearly demonstrates the similarity of these concepts, but also suggest that these measures are not redundant.

For females the cognitive scale of the BES was more strongly associated with the PT scale than the EC scale of the IRI. For males the cognitive scale of the BES was similarly related to the PT and EC scales \((r = .33 \text{ and } r = .30)\). Interestingly, the affective scale of the BES was more strongly related to the PT scale than the EC scale for males and especially for females.

**Table 1**
Correlation of the BES to theoretically relevant constructs

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th></th>
<th></th>
<th>Females</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BES</td>
<td>BES</td>
<td>BES</td>
<td>BES</td>
<td>BES</td>
<td>BES</td>
</tr>
<tr>
<td></td>
<td>cognitive</td>
<td>affective</td>
<td>total</td>
<td>cognitive</td>
<td>affective</td>
<td>total</td>
</tr>
<tr>
<td>IRI Empathic concern</td>
<td>0.30</td>
<td>0.39</td>
<td>0.41</td>
<td>0.29</td>
<td>0.18</td>
<td>0.27</td>
</tr>
<tr>
<td>IRI Perspective taking</td>
<td>0.35</td>
<td>0.51</td>
<td>0.51</td>
<td>0.37</td>
<td>0.44</td>
<td>0.48</td>
</tr>
<tr>
<td>IRI total</td>
<td>0.36</td>
<td>0.52</td>
<td>0.53</td>
<td>0.38</td>
<td>0.35</td>
<td>0.43</td>
</tr>
<tr>
<td>TAS</td>
<td>-0.21</td>
<td>-0.09*</td>
<td>-0.17</td>
<td>-0.31</td>
<td>-0.07*</td>
<td>-0.20</td>
</tr>
<tr>
<td>Thurstone verbal fluency</td>
<td>0.09*</td>
<td>0.01*</td>
<td>0.06*</td>
<td>0.14</td>
<td>0.15</td>
<td>0.17</td>
</tr>
<tr>
<td>UPPS urgency scale</td>
<td>-0.08*</td>
<td>-0.02*</td>
<td>-0.05*</td>
<td>-0.08*</td>
<td>0.11</td>
<td>0.04*</td>
</tr>
<tr>
<td>Big five</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extraversion</td>
<td>0.16</td>
<td>0.06*</td>
<td>0.13</td>
<td>0.25</td>
<td>-0.04*</td>
<td>0.09*</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>0.26</td>
<td>0.23</td>
<td>0.30</td>
<td>0.23</td>
<td>0.18</td>
<td>0.24</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>0.16</td>
<td>0.13</td>
<td>0.17</td>
<td>0.06*</td>
<td>-0.02*</td>
<td>0.01*</td>
</tr>
<tr>
<td>Openness</td>
<td>0.34</td>
<td>0.24</td>
<td>0.34</td>
<td>0.18</td>
<td>0.09*</td>
<td>0.15</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>-0.10*</td>
<td>0.10*</td>
<td>0.02*</td>
<td>-0.07*</td>
<td>0.28</td>
<td>0.16</td>
</tr>
<tr>
<td>Poor parental supervision</td>
<td>-0.12</td>
<td>-0.20</td>
<td>-0.20</td>
<td>-0.09*</td>
<td>-0.12</td>
<td>-0.12</td>
</tr>
<tr>
<td>Low SES</td>
<td>-0.10*</td>
<td>-0.07*</td>
<td>-0.10*</td>
<td>-0.14</td>
<td>-0.10*</td>
<td>-0.13</td>
</tr>
</tbody>
</table>

\(*-p>.05.*

**Relationship between BES and IRI**

Table 1 shows the Pearson Correlations between the cognitive, affective and total scales of the BES and the theoretically related measures for males and females separately.

The similarity between the total of the BES and the total of the IRI for males and females was reflected in their relatively high inter-correlations \((r = .53 \text{ males}, \ r = .43 \text{ females})\). This clearly demonstrates the similarity of these concepts, but also suggest that these measures are not redundant.

For females the cognitive scale of the BES was more strongly associated with the PT scale than the EC scale of the IRI. For males the cognitive scale of the BES was similarly related to the PT and EC scales \((r = .33 \text{ and } r = .30)\). Interestingly, the affective scale of the BES was more strongly related to the PT scale than the EC scale for males and especially for females.

**Relationship between BES and alexithymia**

A significant negative relationship was evident between the total BES scale and the TAS for both males \((r = -.17)\) and females \((r = -.20)\). However, it can be seen that this outcome is the result of negative relationship between the cognitive scale of the BES and the TAS. The correlations between the affective scale and the TAS were also negative, but non-significant.
Relationship between BES and verbal fluency (intelligence)

When the scales of the BES were compared to the verbal fluency measure significant but small positive correlations were found, but only for females. For females, the cognitive scale correlated $r = .14 \ (p < .0001)$, the affective scale correlated $r = .15 \ (p < .0001)$ and the total scale correlated $r = .17 \ (p < .0001)$. None of the correlations for males were significant.

Relationship between BES and impulsivity

When the scales of the BES were compared to the measure of impulsivity, only one significant relationship was evident; the affective empathy of females had a low but significant positive correlation.

Relationship between BES and personality

For both males and females, a significant positive correlation was evident between the cognitive measure of the BES and extraversion, and this relationship was also significant when extraversion was compared to the total BES scale for males, but not females. As expected, all scales of the BES show significant and moderate positive correlations with agreeableness for both males and females. Interestingly, there were positive correlations between all of the scales of the BES and conscientiousness, but only for males; for females all correlations with this personality dimension were non-significant. Moderate correlations existed between the cognitive, affective and total scales of the BES and openness for males, but the correlations were smaller and were only significant for the cognitive and total scales for females. For males there were no significant relationships between the BES and neuroticism. For females, positive and significant correlations were found between neuroticism and the affective and total scales of the BES.

Relationship of BES to parental supervision and parental SES

It can be seen that poor parental supervision was significantly and negatively associated with all of the scales of the BES for males. For females, poor parental supervision was negatively associated with the affective and total scales of the BES. For both males and females, the relationship between poor parental supervision and low empathy was stronger for affective rather than cognitive empathy.

All scales of the BES were negatively related to low SES in both males and females, but this was only significant for the cognitive and total scales for females.

Relationship between BES and social desirability

In order to assess the relationship between the BES and social desirability those in the second wave of data collection completed a scale composed of six items from the Lie scale of the Eysenck Personality Questionnaire (Eysenck & Eysenck, 1991). For males the correlations between this social desirability measure and the BES cognitive, affective and total scales were $r = -0.03$, $r = 0.03$ and $r = 0.00 \ (all \ n.s.)$, respectively. The figures for the females were $r = -0.02, r = -0.10$.
and \( r = -0.11 \) (all n.s.). This clearly indicates that there was no tendency for elevations in empathy scores to be related to elevations in social desirability.

To place this finding in context the social desirability scale was correlated with the PT and EC scales of the IRI, the measure commonly used to assess empathy in previous studies. The comparable figures for the PT, EC and total scale were \( r = 0.29 \) \((p < .0001)\), \( r = 0.18 \) \((p < .017)\) and \( r = 0.27 \) \((p < .0001)\) for males and \( r = 0.33 \) \((p < .0001)\), \( r = 0.15 \) \((p < .050)\) and \( r = 0.29 \) \((p < .0001)\) for females. This indicates that the IRI is more prone to self-presentation bias.

Males and females did not differ significantly on the measure of social desirability \((t = 0.9, \text{n.s.})\). This suggests that female’s higher scores on the BES were not the result of biased responding.

**Relationship between BES and prosocial behaviour**

In the questionnaire subjects were asked about their responses to witnessing a classmate being bullied. The possible responses were (a) nothing, it’s none of my business, (b) nothing, but I think I ought to try to help, and (c) I try to help him or her in some way. While 29.3% of the sample claimed to help someone when they were being bullied, 43.4% did nothing but thought they ought to and 27.3% thought bullying was none of their business. These figures differed significantly between males and females. The figures for males were 25.6%, 37.4% and 36.9% and the corresponding figures for females were 36.4%, 49.9% and 16.9%. This difference was significant \((\chi^2 = 36.01, 2\text{df}, p < .0001)\). More females helped and more males thought it was none of their business.

Table 2 illustrates the means, standard deviations, significance tests and effect size for the comparisons of those who helped others (H), those who thought they should help (SH), and those who thought bullying was none of their business (NB). For example, males who thought that bullying was none of their business had a mean of 31.3 on the cognitive scale of the BES compared to a mean of 32.8 for those males who reported helping others. This difference was significant \((t = 2.05, p < .042)\). As expected, those who helped others differed significantly from those who thought bullying was none of their business on the cognitive, affective and total scales of the BES for both males and females. The effect sizes for all six comparisons ranged from \( d = .28 \) (males cognitive) to \( d = .42 \) (female total) suggesting a range in the difference in proportions from approximately 14% to 21%.

Similarly, significant differences and moderate effect sizes were evident when those who thought bullying was none of their business were compared to those who thought they ought to help, but this was only the case for males. For example, males who thought bullying was none of their business scored 30.3 on the affective scale compared to a mean of 33.4 for those who thought they ought to help. This difference was highly significant \((t = 4.13, p < .0001)\). The difference between females who thought bullying was none of their business and those who thought they ought to help on the measures of empathy were low and non-significant.

**Discussion**

This paper addressed the development and validation of a self-report measure of empathy. The definition of empathy by Cohen and Strayer (1996) provided the conceptual foundation for the
Factor analysis was used to develop highly internally valid cognitive and affective scales of empathy. Confirmatory factor analysis was used, and reinforced the validity of the cognitive and affective components of the scale.

Generally, the results of the comparisons were in line with previous research and theoretical expectation, and therefore support the validity of the BES. For example, in line with previous empathy research (e.g. Lennon & Eisenberg, 1987), females scored much higher than males on all scales of the BES. In fact these differences were so profound that males and females needed to be analysed separately in all subsequent analyses. Subsequent analysis was able to demonstrate that these empathy differences were true differences and not the result of greater socially desirable responding by females. Future research should attempt to determine what underlies the significantly higher cognitive and affective empathy of females. This is especially relevant as some have suggested that these empathy differences may underlie the great disparity in criminal involvement between males and females (Broidy, Cauffman, Espelage, Mazeroile, & Piquero, 2004).

The validity of the BES was also supported by the significant overlap evident between cognitive and affective empathy as measured by the BES. The magnitude of this relationship \((r = .41\) for males, \(r = .43\) for females ) is indicative of the degree of overlap that would be expected from similar interpersonal abilities or traits, but also highlights the differences which might exist in the underlying composition of cognitive and affective empathy. For example, a high degree of emotional recognition would be expected to facilitate both greater cognitive and affective empathy, but being emotionally labile might only influence ones’ affective empathy.

Similarly, the comparison of the scales of the BES to those of the IRI also resulted in correlations, the magnitudes of which illustrate the similarity of emotional congruence and emotional comprehension to sympathy and a global understanding of another’s perspective while also highlighting their distinctions (e.g. Davis, 1983; Eisenberg & Strayer, 1987). The pattern of the relationships between these constructs suggested that there might be a stronger relationship between experiencing another’s emotions and understanding that person’s perspective compared

---

### Table 2
Comparison of those who helped, those who thought they should help and those who thought it was none of their business

<table>
<thead>
<tr>
<th></th>
<th>H M (S.D.)</th>
<th>SH M (S.D.)</th>
<th>NB M (S.D.)</th>
<th>H vs. NB t (p)</th>
<th>Effect size</th>
<th>SH vs. NB t (p)</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Males</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BES cognitive</td>
<td>32.8 (5.6)</td>
<td>32.6 (4.7)</td>
<td>31.3 (5.3)</td>
<td>2.05 (p &lt; .042)</td>
<td>0.28</td>
<td>2.13 (p &lt; .034)</td>
<td>0.26</td>
</tr>
<tr>
<td>BES affective</td>
<td>32.8 (6.5)</td>
<td>33.4 (6.5)</td>
<td>30.3 (6.1)</td>
<td>2.97 (p &lt; .003)</td>
<td>0.40</td>
<td>4.13 (p &lt; .001)</td>
<td>0.49</td>
</tr>
<tr>
<td>BES total</td>
<td>65.5 (10.4)</td>
<td>66.0 (9.2)</td>
<td>61.5 (9.6)</td>
<td>2.99 (p &lt; .003)</td>
<td>0.40</td>
<td>3.91 (p &lt; .001)</td>
<td>0.48</td>
</tr>
<tr>
<td><strong>Females</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BES cognitive</td>
<td>36.1 (3.6)</td>
<td>34.4 (3.9)</td>
<td>34.6 (4.4)</td>
<td>2.44 (p &lt; .015)</td>
<td>0.39</td>
<td>0.26 (n.s.)</td>
<td>−0.05</td>
</tr>
<tr>
<td>BES affective</td>
<td>41.0 (5.6)</td>
<td>40.2 (5.6)</td>
<td>39.0 (6.4)</td>
<td>2.06 (p &lt; .041)</td>
<td>0.34</td>
<td>1.36 (n.s.)</td>
<td>0.21</td>
</tr>
<tr>
<td>BES total</td>
<td>77.1 (7.9)</td>
<td>74.7 (8.1)</td>
<td>73.6 (8.9)</td>
<td>2.62 (p &lt; .010)</td>
<td>0.42</td>
<td>0.83 (n.s.)</td>
<td>0.13</td>
</tr>
</tbody>
</table>

*Note: H = those who helped, SH = those who thought they should help, NB = those who thought it was none of their business.*
to feeling sympathy for that person. This has intuitive appeal as researchers have suggested that sympathy may be a step removed from empathy (Eisenberg & Strayer, 1987).

A negative relationship was found between cognitive and total empathy and alexithymia for both males and females. This suggests that understanding the emotions of another might be inhibited by deficiencies in the experience or understanding of one’s own emotions. Although the nature of this study cannot establish the causal order, it might be expected that the deficiencies inherent in alexithymia, namely difficulties in the expression and understanding of one’s emotions, would negatively influence one’s cognitive empathy rather than the reverse. Future research should examine whether alexithymia or alexithymic traits may account for some of the cognitive empathy differences which are proposed to exist between those who act antisocially and those who do not.

The relationship between the measure of intelligence and the BES was as expected for females, but not for males. This may have been a function of the intelligence measure used in this study as adolescent females are known to be more verbally proficient than adolescent males (e.g. Chase-Lansdale et al., 1995). It would be useful to compare the BES to more comprehensive measures of intelligence and neuropsychological tests which provide more accurate measures of executive brain functioning to determine the extent to which these abilities might be related to empathy.

Unfortunately, the relationship between impulsivity and the BES was not clearly delineated for either males and females. No identifiable relationship existed between impulsivity and the cognitive, affective or total scales of the BES for males and only for the affective scale for females, but even this was in the unexpected direction (positive correlation). Unfortunately, there is no previous research with which to inform these findings. Perhaps no relationship does exist between impulsivity and empathy, however, it is more likely that a more comprehensive study (e.g. multitrait-multimethod) might allow for the elucidation of this relationship.

The relationships between the scales of the BES and the measure of personality were generally as anticipated. As empathy is widely considered to be a factor in altruistic behaviour (e.g. Batson, Turk, Shaw, & Klein, 1995), it would be expected that agreeableness would correlate positively with all measures of empathy. The imaginative ability represented in openness may account for the moderate correlations with the BES. Being a cognitive skill, imaginative ability may account for the stronger relationships between the cognitive scale and openness as opposed to the affective scale. Although Extraversion is usually associated with sensation-seeking (and therefore might be expected to have a negative relationship with empathy similar to that proposed, but not found for empathy and impulsivity), this factor is also considered a measure of general social skills (e.g. John & Srivastava, 1999). Therefore, the low positive correlations between the BES and extraversion may represent overlapping social abilities. As neuroticism is usually associated with anxiety and irritability, negative correlations between empathy and neuroticism would have been expected. However, a moderate positive correlation was evident between neuroticism and affective empathy for females. This may reflect certain facets of neuroticism, namely self-consciousness (which might facilitate the experience of empathic emotions) and guilt (which might be facilitated by the experience of empathy).

Consistent with previous assertions, poor parental supervision was negatively related to the BES scales of empathy (with the exception of cognitive empathy in females). This suggests that poor parental supervision may contribute to lower empathy in children (especially affective empathy) or alternatively, that low empathy may make children difficult to supervise. If future
longitudinal research implicates the former, and low empathy is identified as a consequence of poor parental supervision, this might provide an interesting mechanism for the well known association between poor parental supervision and subsequent violence (e.g. Farrington, 1998).

Low SES was also negatively related to the BES scales of empathy, but this was only significant for females. This finding is interesting in light of research which has demonstrated that social deprivation may have a greater influence on female offending than male offending (Farrington & Painter, 2004). It is plausible that low SES may increase the likelihood of low empathy in females, which in turn might be related to offending. It is likely that the effects of SES on empathy are mediated by child-rearing methods and parental models.

Unlike the IRI, the BES did not have positive correlations with the measure of social desirability. This suggests that the responses to the BES were not influenced by the adolescent’s desire to appear more empathic than they actually were. Future research should continue to examine possible relationships between the BES and other measures of social desirability. It is important to have a measure of empathy which measures true empathy rather than a desire to appear empathic.

In line with expectation, both males and females who reported helping others being bullied scored significantly higher on empathy than those who thought it was none of their business. This supports the view that those who act in a prosocial manner have higher empathy than those who do not. Interestingly, there were also empathy differences between those who thought they ought to help and those who thought that bullying was none of their business, but this was only true for males. It may be that factors apart from empathy (such as the type of help required, the desire to appear tough, etc.) might have had an influence on helping in males, but not in females. Future research should compare the BES to different types of prosocial behaviour, and the frequency of prosocial behaviour. Empathy may not be related to one-off helping, but to frequent helping.

While a great deal of the discussion of this paper has focussed on the relationship between low empathy to antisocial behaviour, this topic was not addressed in the validation exercise of the BES. This was because research has yet to clearly establish that those who commit offences have low empathy, and therefore antisocial behaviour may not provide a convincing platform in which to establish the validity of the BES. However, recent research comparing low empathy to school bullying has been found to support the validity of the BES for this use (Jolliffe & Farrington, 2006).

Conclusion

This validation exercise demonstrated that the BES Empathy Scale has sufficient construct validity. That is, the BES showed significant differences in the expected direction and magnitude between males and females and between those who helped someone being bullied and those who did not. Furthermore, the BES was found to have convergent validity (convergent in the expected direction and magnitude), as demonstrated by the observed relationships with measures of sympathy, perspective taking, alexithymia, agreeableness, conscientiousness, openness, parental supervision and socioeconomic status. The BES also had divergent validity (divergent in the expected directions and magnitude), as demonstrated by the non-relationship with socially desirable responding.
In summary, it appears that the BES is a valid tool by which to measure empathy. However, future validation research is needed especially with prospective longitudinal studies to examine the predictive validity of the BES. It would be particularly useful to examine the influence of changes in empathy on changes in behaviour over time. For the use of the BES for research purposes, please apply to the first author.

References


