This is the pre-peer reviewed version of the following article: Langdon, Murphy, Clare, and Palmer (2010). The psychometric properties of the Socio-Moral Reflection Measure – Short Form and the Moral Theme Inventory for men with and without intellectual Disabilities. Research in Developmental Disabilities, 31, 1204-1215, which has been published in final form at 10.1016/j.ridd.2010.07.025.

# The psychometric properties of the Socio-Moral Reflection Measure – Short Form and the Moral Theme Inventory for men with and without intellectual disabilities.

Peter E Langdon
University of East Anglia
and
Hertfordshire Partnership NHS Foundation Trust

Glynis H Murphy
University of Kent

Isabel CH Clare
University of Cambridge

Emma J Palmer
University of Leicester

## Author Note

Peter E. Langdon, School of Medicine, Health Policy and Practice, University of East Anglia, UK and Broadland Clinic, Hertfordshire Partnership NHS Foundation Trust, UK; Glynis H. Murphy, Tizard Centre, University of Kent, UK; Isabel C.H. Clare, Cambridge Intellectual & Developmental Disabilities Research Group, Department of Psychiatry, University of Cambridge, UK; Emma J. Palmer, School of Psychology, University of Leicester, UK.

This research was supported by a Personal Award Scheme Researcher Development Award from the National Institute for Health Research awarded to Peter E. Langdon. Additional support was provided by The Co-operative Group.
Isabel Clare is funded by the National Institute for Health Research's CLAHRC for Cambridgeshire & Peterborough.

We would like to thank the men who participated in this study.

Correspondence concerning this article should be addressed to Dr Peter E Langdon, School of Medicine, Health Policy and Practice, Faculty of Health, University of East Anglia, Norwich, NR4 7TJ, United Kingdom. Email: P.Langdon@uea.ac.uk

**Abstract**

Drawing conclusions from the literature regarding the moral development of people with intellectual disabilities (IDs) is difficult because of the use of unstandardised and idiosyncratic measures. In order to address this short-coming, a moral reasoning production measure (the Sociomoral Reflection Measure-Short Form; SRM-SF) and a recognition measure (the Moral Theme Inventory; MTI) were presented to men with and without IDs who had no known history of engaging in illegal behaviour. The instruments were completed on two occasions, separated by a two week interval, in order to investigate their basic psychometric properties. The results indicated that there was a strong relationship between the MTI and the SRM-SF, suggesting that the MTI has convergent validity. The internal consistency of the MTI and the SRM-SF ranged from moderate to substantial for both men with and without IDs. However, the test-retest reliability of the MTI was poor for men with IDs, while it was good for men without IDs. The test-retest reliability of the SRM-SF was good for both men with and without IDs. Comparison of the moral reasoning abilities of men with and without IDs suggested that many of the differences between the two groups could be accounted for by general intellectual functioning. The exception was overall score on the SRM-SF and moral reasoning in relation to the law, where men with IDs scored at stage 2(1), when intelligence was controlled. The results were interpreted by suggesting that the relationship between moral reasoning and illegal behaviour may take an inverted U curve shape, moderated by intelligence.

*Keywords*: Moral Reasoning; Moral Development, Intellectual Disability; Cognitive Development; Behavior; Learning Disability; Developmental Disability, Sociomoral Reasoning; Moral Judgement

# The psychometric properties of the Socio-Moral Reflection Measure – Short Form and the Moral Theme Inventory for men with and without intellectual disabilities.

## 1.0 Introduction

Following the early work of Piaget (1932) on the moral development of children, Kohlberg (1969, 1976) revised Piagetian perspectives, to incorporate the moral development that occurs in adolescence and adulthood. He proposed a stage theory of moral development that extended beyond childhood and into adolescence and adulthood. The theory originally comprised six stages, spread across three levels, and formed a hierarchical stage model where more complex levels of moral reasoning required successful progression through earlier stages in a more or less linear fashion. The progression in moral reasoning was accompanied by a parallel developmental progression in logical reasoning. However, Kohlbergian moral development theory has been widely criticised (Gilligan, 1982; Schweder, 1982; Sullivan, 1977) and has subsequently been revised into a sociomoral stage theory (Gibbs, 1979, 2003, 2010). Gibbs (1979) removed post-conventional moral reasoning from Kohlbergian theory arguing that such mature levels were "existential", citing evidence that post-conventional moral reasoning is achieved infrequently across cultures. Gibbs and his colleagues (1979; Gibbs, Basinger, & Fuller, 1992) proposed a sociomoral stage theory (Table 1) regarding the reasons or justifications people give for their behaviour, and these revisions have been shown to have cross-cultural validity (Gibbs, Basinger, Grime, & Snarey, 2007). However, Kohlbergian and Gibbian approaches to moral development are nested within the cognitive developmental domain, and others have adopted alternative theoretical approaches to moral development, nested within the social domain (Semetana, 1999;

Turiel, 1983, 2002) or the emotional domain (Eisenberg, Reykowski, & Staub, 1989; Hoffman, 2000).

INSERT Table 1 About Here

While there are shared commonalities between differing theoretical approaches to moral development, none has actively considered the moral development of people with intellectual disabilities (IDs). Langdon et al. (in press-a) reviewed the literature relating to the moral development of people with IDs, and concluded that the moral development of children, adolescents and adults with IDs appears to be similar to that of their typically developing peers, but occurs in a slower manner. However, the differences in the rate of development between people with IDs and typically developing individuals may disappear when cognitive ability is controlled. However, Langdon et al. (in press-a) suggested that any conclusions must be tenuous because, first, existing studies have not considered the impact of language ability and performance on measures of moral reasoning and, secondly, many of the moral reasoning measures used within the studies are idiosyncratic and unstandardised.

Langdon et al. (in press-a; in press-c) went on to discuss the methods that are traditionally used to measure moral reasoning. Moral reasoning measures are generally classed into two types, a) recognition, and b) production instruments. Recognition instruments involve the presentation of a set of moral justifications to people preceded by the presentation of a moral dilemma. Respondents are asked to choose justifications which best match their own moral reasoning about the dilemma. Production instruments

are different because participants are asked to verbalise their own reasoning in response to questions which follow the reading of a moral dilemma. Langdon et al. (in press-a) suggested that, since people with IDs may have communication difficulties, recognition instruments may have greater utility with this population, because they may be easier to understand. However, many recognition instruments do not measure the developmentally younger stages of moral reasoning (Rest, 1979; Rest, Narvaez, Thoma, & Bebeau, 1999) and their validity is questionable, because in contrast to production instruments, they often do not discriminate reliably between offender and non-offender participants. (Basinger & Gibbs, 1987; Gavaghan, Arnold, & Gibbs, 1983; Gibbs et al., 1984; Stams et al., 2006).

Considering this measurement problem further, Langdon et al. (in press-a) also suggested that the Socio-Moral Reflection Measure- Short Form (SRM-SF) may be used with people with IDs, but concluded that further research was required. The SRM-SF is a measure that is linked to Gibb's Cognitive-Developmental Model of Sociomoral Reasoning (Gibbs, 2003, 2010). As Langdon et al., (in press-a) point out, this measure has been successfully used with young children (Gibbs et al., 2007), and no reading or writing is required if the instrument is presented as part of an interview. However, as yet there is no evidence that recognition or production instruments can be reliably used to assess the moral reasoning abilities of people with IDs.

As a consequence of the issues raised by Langdon et al. (in press-a; in press-b), the aims of this study were twofold. First, the study sought to examine the psychometric

properties of a production and recognition instrument of moral reasoning in relation to men with and without IDs. Secondly, the study aimed to compare the moral reasoning abilities of men with and without IDs using both forms of measurement, controlling for language ability or intelligence. Men with and without IDs were recruited from the community in the Eastern region of the United Kingdom and completed a battery of assessment measures in a single session. Two weeks later, the participants completed the measures of moral reasoning again so that the test-retest reliability of the instruments could be examined. Assuming that recognition instruments should be easier to understand, it was hypothesised, compared to the production measure (the Socio-Moral Reflection Measure – Short Form), the recognition measure (the Moral Theme Inventory) would possess superior psychometric properties when used with men with IDs. It was also hypothesised that men with IDs would have developmentally earlier moral reasoning abilities, and differences from men without IDs would be partially accounted for by language ability, and fully accounted for by intelligence.

## 2.0 Methods

### 2.1 Participants

Thirty two men (*M* age=45.88, SD=15.01; *M* Full Scale IQ=59.35, SD=6.16) were recruited from services for people with IDs in the eastern region of the United Kingdom and formed the *IDs Group*. Twenty-eight men (*M* age=40.64, SD=10.41; *M* Full Scale IQ=102.29, SD=8.05) without IDs were also recruited from community sources and formed the *Comparison Group*. All of the participants included in this study reported that they were of white British ethnic origin. All of the participants with IDs

had attended a special school for people with intellectual or other developmental disabilities, and were now using adult intellectual disability services.

The specific inclusion criteria were: a) all participants should be men because there is some evidence that men and women make moral judgements differently (Gilligan, 1982; Kohlberg, Levine, & Hewer, 1983, 1984; Walker, 1995), b) the Full Scale IQ of participants with IDs should be less than 70, with associated difficulties with adaptive behaviour (considered to be present if the person was receiving support from specialist services for health and social care), and these difficulties having an onset before the age of 18 (American Psychiatric Association, 2000), and c) the Full Scale IQ of participants without IDs should be greater than 70, and there should be no associated difficulties with adaptive behaviour. Adaptive behaviour difficulties were assumed to be absent if the person was employed and was not receiving support from specialist services. Participants were excluded if they had a known history of criminal charges, cautions or convictions or they were currently subject to criminal justice proceedings, including ongoing criminal investigations. The exclusion of these participants was based on the relationship between moral reasoning and illegal behaviour (Blasi, 1980; Nelson, Smith, & Dodd, 1990; Stams et al., 2006) amongst young offenders. Finally, participants with IDs were excluded if they were judged to lack the capacity to provide consent to take part in this research.

**2.1.1Attrition.** There was some attrition associated with the current study. Among the IDs Group, four participants withdrew from the study at differing points

during the assessment process, and one further participant was withdrawn by the researcher because he appeared overly anxious while completing some of the material. Of the four participants who withdrew, one telephoned the researcher and left a message stating that he no longer wanted to take part. Another stated that he did not want to take part further because of competing activities at his day centre. When asked if he would like to rearrange, he declined. The two other participants asked to stop once the assessment had begun and no longer wanted to take part.

Among the Comparison Group, none of the participants withdrew from the study, or were withdrawn by the researcher. However, one participant did not return some of the questionnaires, and reminders did not result in the return of his data; another participant did not return one questionnaire, and again, reminders did not result in its return.

**2.2 Design and Procedure**

Two groups of participants (Group: IDs or Comparison) were recruited and completed a set of measures at one time point, and then completed some of the measures again, following a two-week interval. This two week time interval allowed for the examination of the test-retest reliability of the moral reasoning instruments.

Following a favourable ethical opinion from the Suffolk NHS Research Ethics Committee, information about the project was disseminated to men with IDs by distributing a poster and a leaflet to intellectual disabilities services in the eastern region

of the United Kingdom. Managers of day services and community learning disabilities teams were contacted directly, and informed of the project. They were asked to distribute information leaflets to men with IDs using their services. They were specifically directed not to share information regarding the study with anyone using their service whom they knew to have a history of engaging in illegal behaviour. Any man who expressed an interest in taking part was asked to alert his key-worker, who then informed the manager. The manager then contacted the researcher to inform him of the number of possible participants at a site, and a mutually convenient time was arranged to attend the site and speak to potential participants. Once someone indicated that he might like to take part, full information about the study was provided, and he was asked to provide signed consent.

Information about the study was disseminated to the Comparison Group in several different ways. Leaflets and information sheets were distributed by their managers to men employed within a university in a non-academic position. Information about the study was also disseminated using an advertisement email system at this university. Participants were asked not to volunteer for the study if they had a history of engaging in illegal behaviour. Interested participants were invited to contact the researcher directly, and signed consent was given by those who wished to take part.

All participants were interviewed on two occasions. During the first meeting, all were asked whether or not they had a history of police arrest or caution, or a criminal conviction, or if they were part of an ongoing trial or police investigation as a defendant

or suspect. Any participant who disclosed such a history was not recruited into the current study.

Initially, the Wechsler Adult Intelligence Scale - III$^{UK}$ (WAIS-III; Wechsler, 1998) was administered to assess the general intellectual functioning of participants, while the spoken language portion of the Test of Adolescent and Adult Language Fourth Edition (TOAL-4; Hammill, Brown, Larsen, & Wiederholt, 2007) was used to assess the spoken language of participants. A measure to assess socio-economic status was also presented. Measures of socially desirable responding and two measures of moral reasoning were presented in a randomised order. For the IDs Group, the assessment material was presented orally. In order to determine test-restest reliability, the measures of moral reasoning were presented following a two-week interval. The measures of socially desirable responding were also administered a second time for a similar purpose, but this was part of another study (Langdon, Clare, & Murphy, in press-b). All participants were paid twenty pounds in shopping vouchers as a token of appreciation for taking part.

## 2.3 Measures

**2.3.1 General Intellectual Functioning.** The Wechsler Adult Intelligence Scale - III$^{UK}$ (WAIS-III; Wechsler, 1998) was used to assess the general intellectual functioning of participants. The WAIS-III is a well developed reliable and valid measure of general intelligence that has been standardised on a British population. Reliability coefficients for the WAIS-III IQ scales range from 0.88 to 0.97 (Tulskey, Zhu, & Ledbetter, 1997).

The WAIS-III yields three different IQ scores. These are called the Verbal IQ, Performance IQ and Full Scale IQ. Verbal IQ reflects acquired knowledge, verbal reasoning and comprehension of information presented within the verbal domain. Performance IQ reflects non-verbal reasoning, visual-spatial processing, attentiveness to detail and visual-motor integration. Full Scale IQ is an aggregate of the Verbal and Performance IQ scores and represents global intellectual functioning.

**2.3.2 Spoken Language.** The Test of Adolescent and Adult Language Fourth Edition (TOAL-4; Hammill et al., 2007) was used to assess the spoken language of participants. The TOAL-4 is a standardised reliable and valid assessment of spoken and written language which assesses semantics, grammar and graphology. It comprises six subtests: three of which assess spoken language and three which assess written language. For the purposes of this study, there was no need to administer the written language subtests because expressive language was of most interest and more likely to be related to moral reasoning scores, especially in relation to production measures. Additionally, many of the items associated with the written language subtests are complex and require writing, and would be difficult for people with IDs to complete.

The three spoken language subtests administered were Word Opposites, Word Derivations, and Spoken Analogies. During Word Opposites, the respondent is asked to say an opposite word to the word spoken by the examiner, while Word Deviations asks the respondent to change a given word so that it is said correctly at the end of a sentence. Spoken Analogies involves the examiner giving the respondent an analogous sentence

which needs to be completed. Scaled scores from the three subtests combine to form an Index of Spoken Language ability.

Test-retest reliability for the TOAL-4 has been reported to range from 0.83 to 0.97 (Hammill et al., 2007) and the test is considered to possess content validity, criterion-prediction validity, and construct-identification validity (Hammill et al., 2007). The measure has been shown to assess language abilities accurately in a number of groups (Hammill et al., 2007).

**2.3.3 Socio-Economic Status.** Measures used as part of the Department for Work and Pensions' Families and Children Study (FACS; Department for Work and Pensions, 2002) were considered because they have subjected to a secondary analysis examining the socio-economic position of families with children and adolescents with IDs (Emerson & Hatton, 2007). Within the original FACS study, socioeconomic status and social capital were assessed across areas such as household income, material and social hardship, household occupation, and debt and savings (Department for Work and Pensions, 2002; Emerson & Hatton, 2007).

However, the purpose of the current study was not to replicate the findings of the FACS study, and not all the areas that were assessed as part of the original study were relevant. Given that the current study included adults with IDs, many of whom were living in supported or residential accommodation and not employed, examining household income and household occupation was problematic. Additionally, trying to

gain access to information relating to debt and savings for participants was ethically problematic.

Nevertheless, one of the aspects that formed part of the FACS study that could be feasibly used, was not overly intrusive, and appeared likely to return useable data, was material and social hardship. This construct was assessed by a short questionnaire (the Hardship and Deprivation Scale; HDS) comprising twenty-eight items that aims to examine what a person would like to buy but is unable to afford, across several domains, including food, clothing, material possessions, and social activities. A total score is calculated by dividing the number of items that the person cannot afford by the sum of the number of items that he currently has and those he cannot afford, ignoring the number of items that he neither wants nor needs (e.g. toys and sports gear for children). The total score ranges from 0 to 100, with higher scores reflecting higher material and social hardship, or lower socio-economic status.

There are no validity or reliability data relating to the Hardship and Deprivation Scale (HDS). However, Emerson and Hatton (2007) used this questionnaire within their study examining poverty and socio-economic status within families where there is a child or an adolescent with IDs. Their demonstrated that 31% of the relationship between child health and IDs can be explained by socio-economic factors, suggesting that the questionnaire may be valid for the assessment of aspects of socio-economic status.

**2.3.4 Social Desirability.** This was used in addition to the assessment of moral reasoning, as there may be a relationship between moral reasoning and social desirability

for two reasons, 1) participants may distort their responses in an attempt to present themselves in a favourable manner or 2) those endorsing a virtuous viewpoint, may, as a consequence, score higher on a measure of social desirability. Langdon et al. (in press-b) revised the Self- and Other-Deception Questionnaires, creating the Self- and Other-Deception Questionnaires – Intellectual Disabilities (SDQ-ID and ODQ-ID) in order to examine the psychometric properties of these instruments for the sample of participants included within the current study. The results indicated that the instruments had moderate to substantial internal consistency and moderate to excellent test-retest reliability in both groups of men with and without IDs. Langdon et al., (in press-b) also demonstrated that differences between men with and without IDs on the SDQ-IQ and the ODQ-ID could be accounted for by differences in intelligence.

**2.3.5 The Moral Theme Inventory.** The Moral Theme Inventory (MTI; Narvaez, Gleason, Mitchell, & Bentley, 1999) is a recognition measure of moral reasoning. The measure was primarily developed for use with children, but has also been used with populations of adults. Respondents are asked to consider four moral stories which have a moral message, and these are presented by playing a digital audio file to each participant. After this, respondents are asked to engage in a series of tasks to assess their moral reasoning (Table 2).

---

Insert Table 2 About Here

---

The first task examines whether or not the respondent has understood the moral story, while the remaining tasks assess his or her moral reasoning. Performance across the Vignette Rating and Choice, and the Message Rating and Choice tasks are adjusted by subtracting the rating for incorrect answers from the rating for correct answers. Scores are then combined into a Composite Score by adding the total score across the Vignette and Message tasks. Higher scores represent developmentally more mature moral reasoning; however, in contrast with some measures of moral reasoning, the score does not directly relate to a theoretical moral stage.

The MTI appears promising for use with people who have IDs because it considers developmentally younger moral reasoning, and the moral stories are recorded and played to participants, rather than having to be read by them. The internal consistency of the Comprehension task and the Composite Score has been reported to be good (Narvaez et al., 1999), and the MTI differentiates between children of differing ages and adults (Narvaez et al., 1999).

However, the MTI is lengthy and was originally designed to be presented over two sessions. This was potentially problematic for the current study because participants were already being asked to complete a large amount of test material. There is some evidence to suggest that the reliability and validity of the MTI is not affected if only two stories are used, rather than the original four (Narvaez & Bock, 2001). Consequently, it was decided to present only two of the four moral stories.

Unfortunately, the content of the moral stories is not culturally appropriate for use within the United Kingdom. The MTI was developed in the United States, and each of the four stories reflect life there. The four moral stories are titled, a) California, b) Malcolm, c) Jed, and d) Kim. Each title is the name of the main character within each story. The first two stories were not used in the current study because there is some evidence that validity and reliability are not compromised by using only the Jed and Kim stories (Narvaez & Bock, 2001). The Jed story is about a boy who was left to care for his baby sister, but invited to play American football with his friends. Some of the language is not used within the United Kingdom (e.g. diapers, sidewalks, strollers) and revision was required. The Kim story presents a family who are travelling from Detroit to Minneapolis in search of employment. Again, much of the language is not used within the United Kingdom (e.g. freeways, baloney sandwiches, cents and dollars), and so also required revision.

The author of the MTI gave permission to revise the moral stories to reflect UK culture. Revisions were shared with the original author of the MTI. American language was removed and replaced with British words (e.g. catching a football was changed to kicking a football). Names of cities in the United States (e.g Detroit) were replaced with British cities (e.g. Liverpool). No alterations, however, were made to the moral theme of each story. New audio recordings were then prepared using a British English speaker.

**2.3.6 The Sociomoral Reflection Measure- Short Form.** The Sociomoral Reflection Measure (SRM-SF) is a production measure of moral reasoning (Gibbs et al.,

1992) and has been shown to possess high levels of test-retest reliability (r=0.88; Gibbs et al., 1992), and excellent internal consistency (k=0.92; Gibbs et al., 1992). The SRM-SF appears valid as it is correlated with the Moral Judgement Interview, and discriminates between children of differing ages, as well as between 'delinquent' and 'non-delinquent' adolescents (Gibbs et al., 1992).

The SRM-SF comprises eleven questions, and generally takes about twenty minutes to administer. The questions relate to the following seven constructs, a) Contract (questions one to three), b) Truth (question four), c) Affiliation (questions five and six), d) Life (questions seven and eight), e) Property (question nine), f) Law (question ten), and g) Legal Justice (question eleven). Each question is relatively brief, and invites the respondent first to consider the importance of behaving in a certain manner, or making a certain decision, within the context of a forced choice. For example, when asked the question, "Think about when you've made a promise to a friend of yours. How important is it for people to keep promises, if they can, to their friends?", the respondent is asked to choose whether this is very important, important, or not important. Next, respondents are asked to consider further by answering the following question, "Why is that very important / important / not important?". Respondents write their answers on the questionnaire, or give them orally to be recorded by the interviewer. All answers from the IDs Group were recorded by the interviewer.

Verbatim answers are scored according to a set of complex rules and heuristics, and the development of proficient and reliable scoring occurs through the use of practice

scoring material (Gibbs et al., 1992). Responses to each question are assigned a developmental rating which corresponds to a moral stage associated with Gibb's Socio-Moral Reasoning Theory. At least seven of the eleven questions must be answered with scorable material in order for a questionnaire to be reliably scored. Once a developmental rating is assigned to each question, it is converted to a number (e.g. a developmental rating of 1 corresponds to moral stage 1, and is assigned the numerical value 1). Scores across all the questions are then summed and the mean is calculated and multiplied by 100, yielding a possible score of 100 to 400. As shown in Table 2, these scores correspond to a person's global moral stage. Additionally, moral stage ratings can be generated for the seven constructs examined by the SRM-SF: a) Contract, b) Truth, c) Affiliation, d) Life, e) Property, f) Law, and g) Justice. The scores generated across these constructs are interpreted using Table 3.

---

Insert Table 3 About Here

---

Since there are no known reliability data regarding the SRM-SF or the MTI for men with IDs, the test-retest reliability and the internal consistency of these measures were examined within the current study. The inter-rater reliability of the scoring of the SRM-SF was also calculated using an expert rater who scored a random sample of 14% (n=15) of completed questionnaires.

**2.4 Data Preparation and Analysis**

All data were entered and analysed using PASW Statistics Version 18.0.2. Descriptive data were generated and examined, and any possible errors were checked and corrected as appropriate. Data were inspected for departures from normality by visual inspection of histograms and the generation of P-P plots. No variables departed substantially from normality with the exception of the TOAL-4 Spoken Language Index, the HDS, the Comprehension section of the MTI, and the seven constructs assessed using the SRM-SF; however, the overall SRM-SF score was not affected. Non-parametric statistics were use for analyses relating to the HDS and the Comprehension section of the MTI.

The statistical analyses were planned and allowed for the effective investigation of the hypotheses. Initially, the IDs Group and the Comparison Group were compared on basic descriptive data relating to age, IQ, socioeconomic status and language. Following this, the psychometric properties of the moral reasoning measures were examined at both Time 1 and Time 2. Internal consistency was determined by calculating a Cronbach's alpha, and two-week test-retest reliability was determined by calculating the intraclass correlation coefficient (single measure reliability). Results were interpreted according to the recommendations of Landis and Koch (1977), Cicchetti and Sparrow (1981) and McDowell (2006).

The relationships between measures of moral reasoning and social desirability, socioeconomic status, age and intelligence and language ability were examined by calculating Pearson correlation coefficients (two tailed) or in relation to the HDS and Spoken Language, Spearman correlation coefficients (two tailed), using Time 1 data

only. Full Scale IQ and Spoken Language were entered as covariates in further analyses. A series of univariate analyses of variance (ANOVA) and analyses of covariance (ANCOVA), controlling for spoken language ability or intelligence were used with bootstrapping; sampling 5000 times with replacement. Bootstrapping is a powerful alternative to parametric statistics and generates robust estimates of standard error and confidence intervals. Using ANOVA and ANCOVA, parameters were estimated and bias corrected and accelerated ($BC_a$) 95% confidence intervals were generated and reported regarding any differences between the groups. In reporting our statistics, the statistic reported is based on the original data, while the significance level is derived using bootstrapping. We also report the BCa confidence interval; if this does not include zero, then differences are statistically significant ($p<0.05$). These analyses only included data collected at Time 1. There were three reasons for this decision: a) the Time 2 data were only included within the study to examine the test-retest reliability of the MTI and the SRM-SF, b) the SRM-SF can be successfully scored when participants have provided scorable answers to at least seven of the questions. Unscorable answers should be treated as missing data, and as a consequence, participants with missing data are excluded from more complex statistical analyses (e.g. MANOVA) that included the factor Time, thus reducing the sample size, and c) there was no significant difference across Time in relation to the MTI ($t(52)=1.15$, $p=0.25$; $BC_a$ 95% CI=-1.17 to 4.46), or the SRM-SF ($t(52)=-1.86$, $p=0.07$; $BC_a$ 95% CI=-10.10 to -0.015).

## 3.0 Results

There were no significant differences between the groups in terms of age ($t(57)=1.48$, $p=0.14$; $BC_a$ 95% CI=-1.79 to 11.78), while the Comparison Group had a significantly higher Full Scale IQ ($t(57)=41.37$, $p<0.001$; $BC_a$ 95% CI=-46.43 to -39.25), Verbal IQ ($t(57)=38.10$, $p<0.001$; $BC_a$ 95% CI=-42.05 to -34.06), and Performance IQ ($t(57)=41.37$, $p<0.001$; $BC_a$ 95% CI=-45.33 to -37.28) as well as a significantly higher Spoken Language score ($t(57)=12.61$, $p<0.001$; $BC_a$ 95% CI=-44.80 to -32.19). The IDs Group scored significantly higher on the Hardship and Deprivation Scale ($z=5.10$, $p<0.001$) indicating that they were of significantly lower socioeconomic status. They also scored significantly higher on the ODQ-ID ($t(57)=5.76$, $p<0.001$; $BC_a$ 95% CI=3.72 to 7.74) and the SDQ-ID ($t(57)=3.43$, $p<0.001$; $BC_a$ 95% CI=1.73 to 5.08) indicating higher levels of social desirability (Table 4).

---

Insert Table 4 About Here

---

**3.1 Psychometric Properties: The Moral Theme Inventory.** The internal consistency of the Vignette Rating Task and the Message Rating Task within the Moral Theme Inventory was examined. The internal consistency of the other tasks associated with this instrument was not examined because the data generated are associated with a single choice, rather than multiple choices across items.

For the IDs Group, the internal consistency of the Vignette Rating Score was substantial at Time 1 ($k=0.80$) and excellent at Time 2 ($k=0.81$); the internal consistency

of the Message Rating Score at Time 1 (k=0.81) and Time 2 (k=0.84; Table 4) was also excellent for the IDs Group. The internal consistency of the Vignette Rating Score at Time 1 was moderate (k=0.46), while at Time 2 it was substantial (k=0.76) for the Comparison Group. The internal consistency of the Message Rating Score was excellent at Time 1 (k=0.81) and substantial at Time 2 (k=0.80). Combining both groups, the internal consistency of the Vignette Rating Score at Time 1 (k=0.67) and Time 2 (k=0.76) was substantial. Turning to the test-retest reliability for the MTI, this was poor ($r_i$=0.20) for the IDs Group, but it was good ($r_i$=0.70) for the Comparison Group. Combining the two groups indicated excellent ($r_i$ =0.81) test-retest reliability (Table 4).

There are no known validity data relating to the MTI. Correlations between the MTI and the SRM-SF, a well-established measure of moral reasoning, were calculated combining both groups of participants. At both Time 1 ($r(55)=0.73$, $p<0.001$) and Time 2 ($r(53)=0.83$, $p<0.001$), there was a significant positive correlation between the MTI and the SRM-SF, indicating that the MTI has convergent validity with respect to the assessment of moral reasoning.

---

Insert Table 4 About Here

---

**3.2 Psychometric Properties: Sociomoral Reflection Measure – Short Form.**

The SRM-SF had substantial internal consistency at Time 1 (k=0.79) and at Time 2 (k=0.67; Table 4) for the IDs Group, while the Comparison Group had substantial

internal consistency at both Time 1 (k=0.77) and Time 2 (0.78). The test-retest reliability of the SRM-SF for the IDs Group was good ($r_i$=0.74), and was also good ($r_i$=0.78) for the Comparison Group. Combining both groups revealed that the internal consistency of the SRM-SF at Time 1 (k=0.95) and Time 2 (k=0.96) was excellent, as was the test-retest reliability of SRM-SF ($r_i$=0.96; Table 4). The inter-rater reliability of the SRM-SF was also examined with respect to both the total score. The results indicated excellent agreement between raters with respect to total score ($r_i$=0.98).

**3.3 Correlations.** The relationship between variables was explored further by examining the correlation coefficients (two-tailed) between variables for the IDs Group and the Comparison Group separately using Time 1 data only. This was carried out to determine which variables would be most appropriate to act as covariates in further analyses. Amongst the IDs Group, there was a significant positive correlation between SRM-SF scores at Time 1 and Full Scale IQ (r(28)=0.44, p=0.018) and spoken language ability (r(28)=0.53, p=0.003). However, there were no significant relationships between SRM-SF scores at Time 1 and age (r(28)=-0.05, p=0.81), socioeconomic status (r(28)=-0.08, p=0.69), the ODQ-ID (r(28)=0.26, p=0.19), or the SDQ-ID (r=-0.20, p=0.32). There was no significant relationship between MTI scores and age (r(28)=-0.09, p=0.63), Full Scale IQ (r(28)=0.18, p=0.36), socioeconomic status (r(28)=-0.08, p=0.67), or social desirability as measured by the ODQ-ID (r(28)=-0.16, p=0.43) and SDQ-ID (r(28)=-0.12, p=0.56). However, there was a significant positive relationship between scores on the MTI at Time 1 and spoken language ability (r(28)=0.40, p=0.036).

Examining the relationships between these variables amongst the Comparison Group, there was a significant positive correlation between SRM-SF scores and Full Scale IQ ($r(27)=0.44$, $p=0.036$) and spoken language ability ($r(27)=0.41$, $p=0.034$). There was no significant relationship between SRM-SF scores and age ($r(27)=-0.05$, $p=0.79$) or socioeconomic status ($r(27)=-0.07$, $p=0.73$). There was a significant relationship between social desirability, as measured by the ODQ-ID ($r(27)=0.51$, $p=0.006$) and the SRM-SF. There was no relationship between the SRM-SF and the SDQ-ID ($r(27)=0.18$, $p=0.37$). There were significant positive relationships between MTI scores and age ($r(28)=-0.49$, $p=0.009$), Full Scale IQ ($r(28)=0.61$, $p=0.001$), and spoken language ability ($r(28)=0.44$, $p=0.024$). In contrast, there were no significant relationships between MTI scores and socioeconomic status ($r(28)=-0.09$, $p=0.64$) or social desirability as measured by the ODQ-ID ($r(28)=-0.19$, $p=0.325$) and SDQ-ID ($r(28)=-0.19$, $p=0.340$).

It was apparent that there was an inconsistent, or absent relationship, between moral reasoning scores and age, socioeconomic status and social desirability across both groups. As a consequence, these variables were not included as covariates in further analyses. Full Scale IQ was related to moral reasoning scores in both groups of men, although MTI scores did not relate to IQ amongst the IDs Group. Spoken language ability was related to moral reasoning scores in both groups. Therefore, it was decided to undertake an initial analysis without controlling for any variables, followed by an analysis controlling for language, and finally, an analysis controlling for intelligence.

**3.4 Moral Reasoning Abilities.** Initially, the moral reasoning abilities of the two groups were compared without including covariates in the analysis. On the SRM-SF, the IDs Group scored significantly lower than the Comparison Group on Contract ($F(1, 54)=75.21$, $p<0.001$; $BC_a$ 95% CI=-130.40 to -84.80), Truth ($F(1, 52)=68.29$, $p<0.001$; $BC_a$ 95% CI=-146.80 to -88.80), Affiliation ($F(1, 54)=93.51$; $BC_a$ 95% CI=-118.81 to -80.20), Life ($F(1, 54)=80.73$, $p<0.001$; $BC_a$ 95% CI=-130.10 to -82.30), Property ($F(1, 50)=69.51$, $p<0.001$; $BC_a$ 95% CI=-175.00 to -108.50), Law ($F(1, 51)=179.02$, $p<0.001$; $BC_a$ 95% CI=-228.20 to -167.80), Legal Justice ($F(1, 48)=137.86$, $p<0.001$; $BC_a$ 95% CI=-189.70 to -136.10), and Total Score ($F(1, 54)=225.20$; $BC_a$ 95% CI=-136.14 to -105.99; Table 4). Overall, the SRM-SF mean score for the IDs Group fell at Stage 2, while it fell at Stage 3 for the Comparison Group (Table 4).

On the MTI, the IDs Group scored significantly lower than the Comparison Group on the Comprehension task ($z=-6.11$, $p<0.001$) indicating that men with IDs had some difficulty understanding and recalling information about the moral stories. The IDs Group also scored significantly lower on the Vignette Rating task ($F(1, 55)=42.82$, $p<0.001$; $BC_a$ 95% CI=-11.90 to -6.33) indicating that they had some difficulty recognising the similarity between four short vignettes and the main story with respect to their moral theme. The IDs Group also had more difficultly recognising the short vignette that actually matched the main moral story ($\chi^2(2)=12.46$, $p=0.002$). On the Message Rating task, the IDs Group had a significantly lower score ($F(1, 55)=43.49$, $p<0.001$; $BC_a$ 95% CI=-15.63 to -8.40), indicating that they were less able to recognise the moral message within the moral story. There were similar findings for the Message

Choice task, where the IDs Group had greater difficulty choosing two messages that matched the moral message within the moral story ($\chi^2(4)=35.91$, $p<0.001$). Overall, the IDs Group scored significantly lower than the Comparison Group with regard to Total Composite Score on the MTI ($F(1, 55)=83.50$, $p<0.001$; $BC_a$ 95% CI=-28.74 to -18.40; Table 4).

When this analysis was completed, controlling for spoken language ability, the results did not change. The IDs Group still scored significantly lower than the Comparison Group on Contract ($F(1, 54)=9.36$, $p<0.001$; $BC_a$ 95% CI=-114.00 to -41.10), Truth ($F(1, 52)=8.38$, $p=0.003$; $BC_a$ 95% CI=-138.30 to -35.90); Affiliation ($F(1, 54)=7.19$, $p=0.025$; $BC_a$ 95% CI=-93.60 to -10.90), Life ($F(1, 54)=11.29$, $p<0.001$; $BC_a$ 95% CI=-114.40 to -36.70), Property ($F(1, 50)=14.21$, $p=0.001$; $BC_a$ 95% CI=-201.30 to -70.50), Law ($F(1, 51)=27.69$, $p<0.001$; $BC_a$ 95% CI=-248.20 to -126.80), and Legal Justice ($F(1, 48)=28.44$, $p=0.001$; $BC_a$ 95% CI=-219.60 to -86.50). When spoken language was controlled, the significant difference between the Groups on the Total SRM-SF Score ($F(1, 54)=35.69$, $p<0.001$; $BC_a$ 95% CI=-124.61 to -63.84; Figure 1) remained.

When spoken language was controlled, a significant difference between the Groups on the Vignette Rating Task ($F(1, 55)=4.86$, $p=0.012$; $BC_a$ 95% CI=-10.13 to -0.96) of the MTI remained, while the Message Rating task ($F(1, 55)=2.46$, $p=0.21$; $BC_a$ 95% CI=-14.74 to 1.62) was no longer significantly different. However, overall the

significant difference between the two groups on the MTI Composite Score (F(1, 55)=6.72, p=0.012; $BC_a$ 95% CI=-23.59 to -3.80) remained.

When Full Scale IQ was controlled, there was no significant difference between the Groups on the SRM-SF on six constructs, Contract (F(1, 54)=<1, p=0.303; $BC_a$ 95% CI=-126.10 to 16.50 ), Truth (F(1, 52)=1.05, p=0.196; $BC_a$ 95% CI=-136.00 to 14.90), Affiliation (F(1, 54)=<1, p=0.994; $BC_a$ 95% CI=-65.30 to 66.90), Life (F(1, 54)=<1, p=0.272; $BC_a$ 95% CI=-84.00 to 27.80), Property (F(1, 50)=1.16, p=0.216; $BC_a$ 95% CI=-157.50 to 47.50), and Legal Justice (F(1, 48)=<1, p=0.459; $BC_a$ 95% CI=-110.40 to 31.00). However, the IDs Group scored significantly lower than the Comparison Group on Law (F(1, 51)=9,48, p=0.001; $BC_a$ 95% CI=-252.90 to -91.80). In relation to the Law construct, moral reasoning for the IDs Group fell at Transition Stage 2(1), while for the Comparison Group it fell near Transition Stage 3(4) when IQ was controlled (Figure 1). However, the difference between the Groups on the Total SRM-SF score was just significant (F(1, 54)=2.86, p=0.05; $BC_a$ 95% CI=-93.10 to -8.88) when Full Scale IQ was controlled.

When Full Scale IQ was controlled, there was no significant difference between the two groups on the Vignette Rating task (F(1, 55)=<1, p=0.582; $BC_a$ 95% CI=-5.35 to 10.79), or Message Rating task (F(1, 55)=<1, p=0.873; $BC_a$ 95% CI=-12.54 to 11.89 ) of the MTI. This was also the case in relation to the MTI Composite Score (F(1, 55)=<1, p=0.768; $BC_a$ 95% CI=-12.67 to 14.52).

## 4.0 Discussion

The aim of this study was to examine the basic psychometric properties of a recognition (the MTI) and production (the SRM-SF) measure of moral reasoning used with a sample of men with IDs. It was predicted that the MTI would have superior psychometric properties in comparison to the SRM-SF. It was also predicted that men with IDs would have developmentally younger moral reasoning abilities than the Comparison Group, and that this difference would partially be accounted for by spoken language ability, and fully accounted for by intelligence.

The results demonstrated that the SRM-SF had satisfactory psychometric properties in relation to men with and without IDs, but, unexpectedly, there are some difficulties with the MTI. Although there was a substantial relationship between the MTI and the SRM-SF, suggesting that the MTI is a valid assessment of moral reasoning ability, the test-retest reliability of the MTI for the group of men with IDs was poor. This finding is likely to reflect the complexity of the tasks and the demands it made on working memory. Participants were asked to listen to a moral story, answer a set of questions, understand four more short stories and compare these to the original moral story. They then had to review a set of moral justifications, again bearing in mind the content of the original moral story. Men with IDs scored significantly lower on the MTI Comprehension Task suggesting that they had difficulties even understanding and recalling details from the main moral story. It was noted during the interviews that the men with IDs seemed to be struggling to understand the tasks associated with the MTI, and as a consequence, some of them may have simply been guessing their subsequent

answers. This undermines the suggestion by Langdon et al. (in press-a, in press-c) that recognition instruments may be more appropriate for people with IDs, although it remains the case that, if simpler recognition instruments were available, they could be useful. The findings indicated that the SRM-SF appears to be a promising instrument for use with people with IDs. While its use appeared relatively straightforward, it was noted that some men with IDs found the question, "How important is it for a person to live, even if that person doesn't want to?" difficult to understand. Some repetition of the questions was required in order to encourage participants with IDs to produce sufficiently scorable speech. However, all of the men with IDs provided scorable answers to at least seven of the eleven questions. This meant that no questionnaires were rejected as unscorable.

In contrast with much of the literature (Langdon et al. in press-a), this study used a well developed production and recognition measure of moral reasoning. This means it is possible to generate some meaningful information about the moral reasoning of men with IDs. Men with IDs were generally reasoning at Global Stage 2, and in contrast, men without IDs were generally reasoning at Global Stage 3. However, inconsistent with our prediction, the differences between the two groups were not accounted for when spoken language was controlled, except for the Message Rating Task on the MTI. These results suggest that differences between men with and without IDs on the SRM-SF cannot be accounted for by spoken language ability, while some of the differences on the MTI can be accounted for by differences in spoken language ability. However, the differences between the two groups were also not completely accounted for by intelligence; differences remained regarding the moral justifications given with regard to obeying the

law, and the overall difference between the two groups remained significant. At first glance, this is puzzling because lower moral reasoning tends to be associated with illegal behaviour (Blasi, 1980; Stams et al., 2006), but yet none of the participants in this study had a known history of illegal behaviour. Young people who do not engage in illegal behaviour have moral reasoning in relation to law and legal justice that tends to fall at Stage 3, while their 'delinquent' counterparts tend to use moral reasoning that falls at Stage 2 (Blasi, 1980; Campagna & Harter, 1975; Gavaghan et al., 1983; Gregg, Gibbs, & Basinger, 1994; Stams et al., 2006; Trevethan & Walker, 1989). In the current study, the men with IDs had no known history of illegal behaviour, but still had less well developed moral reasoning abilities.

The explanation may lie in the moral justifications given by men with IDs in relation to the law. In the current study, scores for these justifications fell within Moral Stage 1, and more specifically at the Transition Stage 2(1). This remained the case even when spoken language or intellectual functioning were controlled. Reasoning at this stage is associated with making justifications based on unilateral authority and rules. Some supportive evidence for this suggestion was generated by Richards, Bear, Stewart and Norman (1992) who noted that, in typically developing children, the relationship between moral reasoning and behaviour may be curvilinear, so that the earlier and later moral reasoning stages are associated with fewer behavioural problems. Langdon et al., (in press-c) also suggested that the relationship between moral reasoning and illegal behaviour may be curvilinear, moderated by intellectual ability. The implication is that men with IDs whose moral reasoning about the law is at a developmentally younger stage

may tend to avoid illegal behaviour, because it is 'against the rules', while those who have moral reasoning about the law that falls within Stage 2, will be more likely to have "borderline" intellectual abilities and may therefore be at "risk" of engaging in illegal behaviour. Illegal behaviour should be less prevalent among men of average intellectual ability whose moral reasoning is also developmentally older. Our findings appear consistent with such a model, but since we did not include any men with a history of illegal behaviour in the present study, we could not test it.

The main strength of the current study is that it allows for some conclusions to be drawn about the advantages of a production instrument, the SRM-SF, for the assessment of moral reasoning in men with IDs. However, there were difficulties with unscorable responses on the SRM-SF, meaning that the factor Time had to be excluded from the analysis. We could have limited the effect of this loss of participants from the multivariate analyses if we had used a much larger sample.

While the current study addresses some of the concerns we have raised about the literature on moral reasoning (Langon et al., in press-b), much remains uncertain. For example, we know very little about the longitudinal moral development of children, adolescents and adults with IDs. Only one previous study (Mahaney & Stephens, 1974), which unfortunately had some methodological problems, has attempted to investigate this issue. With some limited exceptions (Moore & Stephens, 1974; Sigman, Ungerer, & Russell, 1983), we also know very little about how, among people with IDs, moral development relates to behaviour. Further investigation is needed into the possibility that

the relationship between moral development and illegal behaviour may approximate an inverted U shape, moderated by intelligence (Langdon et al., in press-a). Nevertheless, since the SRM-SF appears to have satisfactory psychometric properties when used with men with IDs, it is now possible to make use of this measure within such research studies, though some further investigation of the psychometric properties of this measure using both men and women, and adolescents with IDs would be appropriate. It should, though, be borne in mind that the SRM-SF is linked to a specific cognitive-developmental model of moral development (Gibbs, 2003, 2010). There are other theoretical approaches to moral development, such as the social domain theory (Semetana, 1999; Turiel, 1983, 2002), and moral reasoning theory where emotion is given a pivotal role (Eisenberg et al., 1989; Hoffman, 2000). These theoretical approaches have not been considered using participants with IDs, and further research is required.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders-IV-TR*: American Psychiatric Association.

Basinger, K. S., & Gibbs, J. C. (1987). Validation of the sociomoral reflection objective measure-short form. *Psychological Reports, 61*, 139-146.

Blasi, A. (1980). Bridging moral cognition and moral action: A critical review of the literature *Psychological Bulletin, 88*, 1-45.

Campagna, A. F., & Harter, S. (1975). Moral judgment in sociopathic and normal children. *Journal of Personality and Social Psychology, 31*, 199-205.

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86*, 127-137.

DWP. (2002). *Measuring Child Poverty*. London: Department for Work and Pensions.

Eisenberg, N., Reykowski, J., & Staub, E. (1989). *Social and Moral Values: Individual and Social Perspectives*. Hillsdale, NJ: Erlbaum.

Emerson, E., & Hatton, C. (2007). Poverty, socio-economic position, social captial and the health of children and adolescents with intellectual disabilities in Britain: a replication. *Journal of Intellectual Disability Research, 51*, 866-874.

Gavaghan, M. P., Arnold, K. D., & Gibbs, J. C. (1983). Moral judgment in delinquents and nondelinquents - recognition versus production measures. *Journal of Psychology, 114*, 267-274.

Gibbs, J. C. (1979). Kohlberg's moral stage theory. A Piagetian revision. *Human Development, 22*, 89-112.

Gibbs, J. C. (2003). *Moral Development and Reality: Beyond the Theories of Kohlberg and Hoffman*. London: Sage Publications

Gibbs, J. C. (2010). *Moral Development and Reality: Beyond the Theories of Kohlberg and Hoffman*. Boston: Pearson, Allyn and Bacon.

Gibbs, J. C., Arnold, K. D., Morgan, R. L., Schwartz, E. S., Gavaghan, M. P., & Tappan, M. (1984). Construction and validation of a multiple-choice measure of moral reasoning. *Child Development, 55*, 527-536.

Gibbs, J. C., Basinger, K. S., & Fuller, D. (1992). *Moral Maturity: Measuring the Development of Sociomoral Reflection*. Hillsdale, NJ: Lawrence Erlbaum.

Gibbs, J. C., Basinger, K. S., Grime, R. L., & Snarey, J. R. (2007). Moral judgement development across cultures: revisiting Kohlberg's universality claims. *Developmental Review, 27*, 443-500.

Gilligan, C. (1982). *In a Different Voice: Psychological Theory and Women's Development*. Cambridge, MA: Harvard University Press.

Gregg, V., Gibbs, J., C., & Basinger, K., S. (1994). Patterns of developmental delay in moral judgement by male and female delinquents. *Merrill-Palmer Quarterly, 40*, 538-553.

Hammill, D. D., Brown, V. L., Larsen, S. C., & Wiederholt, J. L. (2007). *Test of Adolescent and Adult Language Fourth Edition*. Austin, TX: Pro-Ed.

Hoffman, M. L. (2000). *Empathy and Moral Development: Implications for Caring and Justice* Cambridge: Cambridge University Press.

Kohlberg, L. (1969). Stage and sequence: The cognitive-development approach to socialization. In D. A. Goslin (Ed.), *Handbook of Socialization Theory and Research* (pp. 347-480). Chicago: Rand McNally.

Kohlberg, L. (1976). Moral stages and moralization: The cognitive developmental approach. In T. Lickona (Ed.), *Moral Development and Behavior: Theory, Research and Social Issues* (pp. 31-53). New York: Holt, Rinehart & Winston.

Kohlberg, L., Levine, C., & Hewer, A. (1983). Moral stages: A current formulation and a response to critics. In J. A. Meacham (Ed.), *Contributions to Human Development* (Vol. 10). London: Karger.

Kohlberg, L., Levine, C., & Hewer, A. (1984). Synopses and detailed replies to critics. In *Essays on Moral Development: The Psychology of Moral Development: The Nature and Validity of Moral Stages* (Vol. II, pp. 320-386). San Francisco: Harper & Row.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Langdon, P. E., Clare, I. C. H., & Murphy, G. H. (in press-a). Developing an understanding of the literature relating to the moral development of people with intellectual disabilities. *Developmental Review, doi: 10.1016/j.dr.2010.01.001*.

Langdon, P. E., Clare, I. C. H., & Murphy, G. H. (in press-b). Measuring social desirability amongst men with intellectual disabilities: The psychometric properties of the Self- and Other-Deception Questionnaire-Intellectual Disabilities. *Research in Developmental Disabilities*.

Langdon, P. E., Clare, I. C. H., & Murphy, G. H. (in press-c). Moral reasoning theory and illegal behaviour by adults with intellectual disabilities. *Psychology, Crime and Law*.

Mahaney, E. J. J., & Stephens, B. (1974). Two-year gains in moral judgement by retarded and nonretarded persons. *American Journal of Mental Deficiency, 79*, 134-141.

McDowell, I. (2006). *Measuring Health: A Guide to Rating Scales and Questionnaires* (3rd ed.). New York: Oxford University Press.

Moore, G., & Stephens, B. (1974). Two-year gains in moral conduct by retarded and nonretarded persons. *American Journal of Mental Deficiency, 79*, 147-153.

Narvaez, D., & Bock, T. (2001). *MTI Guide: Guide for Using the Moral Theme Inventory*. Notre Dame, IN: University of Notre Dame.

Narvaez, D., Gleason, T., Mitchell, C., & Bentley, J. (1999). Moral theme comprehension in children. *Journal of Educational Psychology, 91*, 477-487.

Nelson, J. R., Smith, D. J., & Dodd, J. (1990). The moral reasoning of juvenile delinquents: a meta-analysis. *Journal of Abnormal Child Psychology, 18*, 231-239.

Piaget, J. (1932). *The Moral Judgement of the Child*. London: Routledge and Kegan Paul Ltd.

Rest, J. (1979). *Development in Judging Moral Issues*. Minnesota: University of Minnesota Press

Rest, J., Narvaez, D., Thoma, S. J., & Bebeau, M. J. (1999). DIT-2: Devising and testing a new instrument of moral judgement. *Journal of Educational Psychology, 91*(4), 644-659.

Richards, H. C., Bear, G. G., Stewart, A. L., & Norman, A. D. (1992). Moral reasoning and classroom conduct: evidence of a curvilinear relationship. *Merrill-Palmer Quarterly, 38*, 176-190.

Schweder, R. (1982). Review of Lawrence Kohlberg's essays in moral development: The philosophy of moral development. *Contemporary Psychology, 1*, 421-424.

Semetana, J. G. (1999). The role of parents in moral development: A social domain analysis. *Journal of Moral Education, 28*, 311-321.

Sigman, M., Ungerer, J. A., & Russell, A. (1983). Moral judgment in relation to behavioral and cognitive disorders in adolescents. *Journal of Abnormal Child Psychology, 11*, 503-511.

Stams, G. J., Brugman, D., Deković, M., van Rosmalen, L., van der Laan, P., & Gibbs, J. C. (2006). The moral judgement of juvenile delinquents: A meta-analysis. *Journal of Abnormal Child Psychology, 34*, 697-713.

Sullivan, E. V. (1977). A study of Kohlberg's structural theory of moral development. A critique of liberal social science ideology. *Human Development, 20*, 353-376.

Trevethan, S. D., & Walker, L. J. (1989). Hypothetical versus real-life moral reasoning among psychopathic and delinquent youth. *Development and Psychopathology, 1*, 91-103.

Tulskey, D., Zhu, J., & Ledbetter, M. (1997). *WAIS-III/WMS-III: Technical Manual*. London: The Psychological Corporation.

Turiel, E. (1983). *The Development of Social Knowledge: Morality and Convention*. Cambridge: Cambridge University Press.

Turiel, E. (2002). *The Culture of Morality: Social Development, Context and Conflict*. Cambridge: Cambridge University Press.

Walker, L. J. (1995). Sexism in Kohlberg's moral psychology. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Moral Development: An Introduction* (pp. 83-108). Needham Heights, MA: Allyn & Bacon.

Wechsler, D. (1998). *WAIS-III (UK): Administration and Scoring Manual*. London: The Psychological Corporation.

Table 1

Gibbs' Sociomoral Stage Theory (Gibbs et al., 1992)

| *Level and Stage* | ***Description*** |
|---|---|
| *Level 1: Immature* | |
| Stage 1: Unilateral and Physicalistic | Moral justifications are based upon unilateral authority and rule based, or related to punitive consequences of the violation of rules. |
| Stage 2: Exchanging and Instrumental | Moral justifications based upon an understanding that has arisen from social interaction. For example, decisions to help others may be justified because that person may help you in the future. However, justifications remain superficial. |
| *Level 2: Mature* | |
| Stage 3: Mutual and Prosocial | Moral justifications are characterised by further decentration, and are based upon a prosocial understanding of emotional states (e.g. empathy), care and good conduct. |
| Stage 4: Systemic and Standard | Further maturity is indexed by the development of an understanding of the complex social structures in which we live. Justifications are also based upon constructs such as rights, values and character within society. Other justifications may be based upon social justice and responsibility or conscience. |

Table 2

Description of the tasks that form the Moral Theme Inventory (MTI). These tasks were completed following the presentation of a moral story.

| *Task* | ***Description*** |
|---|---|
| 1. Comprehension | Respondents are presented with ten true or false questions which aim to examine their understanding of the moral story. Questions are presented which cover factual information about the dilemma, and inferences that would have to be made from the story. |
| 2. Vignette Rating | Respondents are asked to consider four short vignettes and consider whether or not the moral message of each vignette is similar to the previously presented moral story. Responses are scored on a 5-point Likert scale. Each of the four vignettes were devised such that only one vignette matched the previously presented moral story in terms of moral message, while the remaining three vignettes matched the previously presented moral story only by having the same characters, actions or setting. |
| 3. Vignette Choice | Respondents are asked to choose one of the four vignettes which best matches the previously presented moral story. Frequency data is generated which ranges from 0 to 2, which indicates the number of correct choices. |
| 4. Message Rating | Respondents are asked to rate a series of moral messages according to how well each message matches the original moral theme of the story. Ratings are made along a 5-point Likert scale. |
| 5. Message Choice | Respondents are asked to choose two previously rated moral messages that they think best matches the theme of the moral story. Frequency data is generated which ranges from 0 to 4 which indicate the number of correct choices. |

Table 3

The relationship between scores on the Sociomoral Reflection Measure – Short Form (SRM-SF) and moral stages.

| *Score* | ***Moral Stage*** |
|---|---|
| 100 to 125 | Stage 1 |
| 126 to 149 | Transition Stage 1(2) |
| 150 to 174 | Transition Stage 2(1) |
| 175 to 225 | Stage 2 |
| 226 to 249 | Transition Stage 2(3) |
| 250 to 274 | Transition Stage 3(2) |
| 275 to 325 | Stage 3 |
| 326 to 349 | Transition Stage 3(4) |
| 350 to 374 | Transition Stage 4(3) |
| 375 to 400 | Stage 4 |

Table 4

Descriptive statistics relating to men with and without IDs.

| | **Men with IDs** | | **Men without IDs** | |
|---|---|---|---|---|
| | ***M*** | **SD** | ***M*** | **SD** |
| Full Scale IQ | 59.35*** | 6.16 | 102.29 | 8.05 |
| Verbal IQ | 61.65*** | 6.21 | 99.75 | 8.83 |
| Performance IQ | 63.81*** | 6.27 | 105.18 | 9.36 |
| Spoken Language | 52.91*** | 7.13 | 91.68 | 15.32 |
| Age | 45.88 | 15.01 | 40.64 | 10.41 |
| Socioeconmic Status (HDS) | 25.00*** | 12.83 | 6.64 | 11.73 |
| Self-Deception | 8.97*** | 4.15 | 5.54 | 2.46 |
| Other-Deception | 11.28*** | 4.80 | 5.50 | 3.03 |
| Sociomoral Reflection Measure | | | | |
| Contract | 222.32*** | 56.02 | 329.32 | 31.72 |
| Truth | 201.79*** | 56.90 | 319.60 | 45.41 |
| Affliation | 215.71*** | 45.62 | 314.81 | 27.96 |
| Life | 208.93*** | 46.75 | 315.74 | 41.11 |
| Property | 158.93*** | 54.52 | 302.17 | 68.22 |
| Law | 151.80*** | 31.86 | 352.08 | 71.44 |
| Legal Justice | 166.67*** | 43.85 | 331.82 | 54.65 |
| Total Score | 199.89*** | 32.88 | 338.24 | 48.51 |
| Moral Theme Inventory | | | | |
| Comprehension | 14.46*** | 2.43 | 19.14 | 1.07 |
| Vignette Rating | 1.11*** | 5.37 | 10.17 | 5.00 |
| Message Rating | 4.70*** | 6.43 | 16.73 | 7.20 |
| Total Score | 8.16*** | 8.27 | 31.73 | 10.86 |

| | **Frequency** | | | **Frequency** | | |
|---|---|---|---|---|---|---|
| Vignette Choice[a] | Zero | One | Two | Zero | One | Two |
| | 12 | 10 | 6 | 2 | 9 | 17 |

| | **Frequency** | | | | | **Frequency** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Message Choice[b] | Zero | One | Two | Three | Four | Zero | One | Two | Three | Four |
| | 2 | 12 | 11 | 2 | 1 | 0 | 0 | 3 | 14 | 11 |

***p<0.01*

****p<0.001*

*Note*. [a]Vignette choice is a frequency count of the number of times a participant correctly chooses the vignette that correctly matches the previously presented moral story in terms of moral theme. Two moral stories were presented to participants, so each participant chose a vignette twice. Therefore, the maximum correct frequency count for a participant is 2.

[b]Message choice is a frequency count of the number of times a participant correctly chose the moral message associated with the moral stories. Two moral stories were presented to each participant, and for each moral story there are two correct moral messages. Therefore, the maximum correct frequency count for a participant is 4.

Table 5

Psychometric properties of the Sociomoral Reflection Measure –Short Form and the Moral Theme Inventory for men with and without intellectual disabilities

| | **Men with IDs** | | **Men without IDs** | | **Combined Group** | |
|---|---|---|---|---|---|---|
| | Time One | Time Two | Time One | Time Two | Time One | Time Two |
| Internal Consistency | α= | α = | α = | α = | α = | α = |
| Sociomoral Reflection Measure – Short Form | 0.79 | 0.67 | 0.77 | 0.78 | 0.95 | 0.96 |
| Moral Theme Inventory | | | | | | |
| Vignette Rating Score | 0.80 | 0.81 | 0.46 | 0.78 | 0.67 | 0.76 |
| Message Rating Score | 0.81 | 0.84 | 0.81 | 0.80 | 0.82 | 0.81 |
| Test-Retest Reliability | $r_i$= | | $r_i$= | | $r_i$= | |
| Sociomoral Reflection Measure – Short Form | 0.74 | | 0.78 | | 0.96 | |
| Moral Theme Inventory | 0.20 | | 0.70 | | 0.81 | |

***Figure 1. Adjusted* means (SEM) across the Sociomoral Reflection Measure-Short Form controlling for Full Scale Intelligence Quotient or Spoken Langauage Ability**