INVESTIGATION OF THE MORAL REASONING OF OFFENDING AND NON-OFFENDING ADOLESCENTS USING THE SOCIOMORAL REFLECTION MEASURE-SHORT FORM

CATHERINE MARIA BRUSTEN

Thesis submitted in partial fulfilment of the requirements of the award of Doctor of Philosophy, Oxford Brookes University, April 2003.

ABSTRACT

The body of work presented in this thesis results from an in depth investigation of the moral reasoning of male offending and male and female non-offending adolescents in the UK using the Sociomoral Reflection Measure-Short Form (Gibbs, Basinger & Fuller, 1992). In Chapter 1 a review of the research relevant to field of study was presented. In Chapter 2 findings from initial investigations into the patterns of moral reasoning of male offending and male and female non-offending adolescents were presented. Key aims of the studies presented in Chapter 2 were to replicate previous findings using the measure amongst offenders and non-offenders, in particular to test claims that offenders were morally immature in relation to non-offending peers, and to investigate gender differences on the measure. In line with predictions, the measure discriminated between offenders and non-offenders, with offenders having significantly lower mean scores than non-offenders. No gender differences were observed in the overall sample. However, some difficulties were found in discriminating between younger male offenders and male non-offenders. Tests of the internal characteristics of the scale were also performed. In Chapter 3, subsequent analyses of data were performed using "categorical" methods. This represented a novel means of tackling the data. A key aim was to test the Gilligan (1982) hypothesis that females and males differ in use of Stage 3 "mutual and prosocial" and Stage 4 "systemic and standard" reasoning. Results of the "categorical" analyses apparently corresponded to the Gilligan hypothesis. Patterns of intra-individual variability in moral stage use were also investigated to test claims regarding structural consistency in moral reasoning. It was found that participants were demonstrating intra-individual variability in their patterns of moral reasoning, this presented a challenge to certain aspects of the moral reasoning model. The final chapter summarised key findings to emerge from the thesis, and discussed these in theoretical context.

ACKNOWLEDGEMENTS

There have been so many people who have contributed to this research project that it would not be possible to acknowledge them all by name. For those whose names do not appear here, please accept my apologies: I am truly grateful to all of you. Many thanks go to Roger Lamb, my main supervisor and also to Michael Argyle, my second supervisor, who sadly did not live to see the completion of this work. Their insight, encouragement and support was invaluable. Access to data used in this thesis was supplied as part of a pilot study being carried out by the Citizenship Foundation of London, partially funded by the Youth Justice Board. Help of both organisations is gratefully acknowledged, and particular thanks are due to Don Rowe from the Citizenship Foundation for his help and support throughout the project. Thanks are also due to all the young people who cheerfully volunteered their time to complete the questionnaires, and to the teachers and practitioners in criminal justice settings who facilitated data collection – in particular, Amanda Dickson-Lewis, who went out of her way to help in this respect.

On a personal level I am deeply grateful to many people, these include Michael Swan and Catherine Walter for their love and support, Alison Day and Helen Kara for being really good friends, Auntie "L" for being sympathetic, and the Quint household in Leuven, Belgium for providing kind words of wisdom and cooking some very good meals. I am indebted to the people of Rethimnon, Crete for providing me with a Mediterranean haven of tranquillity. In particular, thanks go to George Mamalakis who provided a Cretan psychologist's perspective on the thorny issue of moral reasoning, thanks are also due to all at Galero and Antico for cheerfully providing coffee and beer and making sure I had a good holiday when needed.

Finally, words are inadequate to express the debt of gratitude that I owe to my husband Luc, who supported me throughout this whole process. His continuing patience, love and cheerful encouragement during some stressful times means more to me than I can say.

CONTENTS

CONTENTS

CONTENTS

CONTENTS

CONTENTS

LIST OF TABLES

LIST OF TABLES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Overview

The findings presented in this thesis result from an in-depth comparative investigation of the moral reasoning of male offending adolescents and male and female non-offending adolescents using the Sociomoral Reflection Measure-Short Form (hereafter referred to as the SRM-SF), carried out amongst a UK based sample. This introductory chapter presents the theoretical background to the measure, reviews key findings in the field and concludes with an overview of the empirical research presented in the current study.

Moral reasoning enhancement forms a core component of many cognitive-behavioural intervention programmes used by practitioners working with young offenders in the UK; however, the assessment of moral reasoning provides certain challenges to the researcher working with young offenders, both in terms of identifying clinical need and in terms of evaluating the relative impact of a moral reasoning intervention on offending behaviour. The first section of this chapter provides a brief historical overview to the use of cognitive-behavioural interventions in the UK and describes some of the issues involved with the assessment of the moral reasoning components of these programmes. The second section of this chapter reviews the theoretical background to the SRM-SF, paying particular attention to Lawrence Kohlberg's cognitive developmental model of moral reasoning. The third section of this chapter provides an outline of the Gibbs et al (1992) revisionist model of moral reasoning, gives a description of the format and content of the measure and reviews the currently published research that has used the SRM-SF. The third section of this chapter also presents some unresolved research issues to be addressed within the current study. The chapter concludes with a description of the body of empirical work to be presented in this thesis.

1.2 Cognitive-Behavioural Interventions with Young Offenders

During the 1970's and 1980's the prevailing view amongst professionals carrying out research into the impact of clinical interventions on the reduction of offending behaviour was that "nothing worked". This pessimism stemmed from a number of wide-ranging surveys carried out in the 1970's

that had failed to uncover evidence to indicate that any programme of intervention could be consistently be relied on to reduce offending behaviour (see McGuire, 1995, McGuire, 2000 for reviews). However, from the mid 1980s researchers started to challenge the prevailing opinion that "nothing worked" (e.g. McGuire & Priestley 1985). Largely as a result of a series of studies using meta-analytic techniques it was demonstrated that the conclusion drawn by researchers that "nothing worked" was flawed, (see e.g. McGuire & Priestley 1995, Vennard et al 1997, McGuire, 2000, Blud et al 2003, for reviews) and that the "net effect" of treatment on offenders actually represented a reduction in recidivism of upwards of 10% (see e.g. Lipsey, 1995). Over the last decade the emphasis in the criminal intervention literature has shifted towards a more positive outlook that has been broadly centred around the so-called "What Works?" (McGuire, 1995) approach.

The most encouraging findings to initially emerge from the "What Works?" literature appeared to concern those programmes that could be broadly categorised as adopting "Cognitive-behavioural" approaches to tackling offending behaviour (see Ross et al, 1988 for a description of "Reasoning and Rehabilitation", which is one of the most widely used programmes to be applied in UK prisons). The majority of cognitive-behavioural interventions are "multi-modal" and focus on addressing a variety of cognitive and behavioural skill deficits that have been identified as criminogenic risk factors amongst the offending population. Typical "core components" of these cognitive behavioural programmes address perceived deficits in social perspective taking, anger management and moral reasoning (Hollin et al, 2002).

A meta analysis carried out by Izzo and Ross (1990) on 46 studies of interventions with young offenders concluded that programmes that included a cognitive component were twice as effective as those that did not. Lipsey (1995) in a meta-analysis of over 400 research studies concluded that well-structured clinical interventions that were focused around "behavioural, training or skills issues" and that maintained a high level of programme integrity were those that were most likely to succeed (Lipsey, 1995:77-78). For a comprehensive review of the initial findings to emerge from the "What Works?" research see McGuire & Priestley (1995); see also Hollin (1990, 1993) for earlier reviews assessing the impact of cognitive-behavioural interventions with young offenders.

Multi-modal cognitive-behavioural intervention programmes are currently the favoured option amongst practitioners in the UK (Vanstone, 2000). However, it has not been until extremely recently that a comprehensive evaluation of the impact of these programmes has been carried out in Britain (see Friendship et al, 2002b, 2003a). The recent evaluation that was carried out by Friendship et al was carried out amongst an adult prison population using reconviction data as an outcome measure. Results of this large scale evaluation have provided some apparently positive findings: when comparing reconviction rates of those who had participated in cognitive behavioural programmes with a matched control sample, it was found that there was a drop in reconviction rates of 14% (Friendship et al 2002b).

McGuire (2000) notes that whilst the evidence regarding the effectiveness of multi-modal cognitive behavioural programmes seems to be encouraging, there is a considerable amount of research that remains to be done in this field (McGuire, 2000). In particular, McGuire considers that assessment of the relative importance of individual components of cognitive programmes needs more detailed analysis and testing. Vennard et al (1997) express some concern at the apparent over-reliance on "reconviction rates" as a source of outcome data, and stress that evaluations of programmes of intervention need to include measurement of change in the target attitudes, behaviour or skills as well as in reconviction rates (Vennard et al 1997). Finally, both McGuire (2000) and Vennard et al (1997) express the need to exercise caution in applying findings obtained from largely North American samples to UK populations.

The majority of cognitive-behavioural intervention programmes that are used with offenders derive from work that has been carried out amongst adults. There are relatively few cognitive-intervention programmes that are specifically designed for use with young offenders. Two programmes that have been developed in the US and are considered to be suitable for use with young adolescent offenders (Hollin et al 2002) are the EQUIP program (Gibbs et al 1995), and Aggression Replacement Training (ART) Programme (Goldstein et al 1998). Both of these programmes adopt a multi-modal approach and incorporate "moral reasoning enhancement" as one of their core components. The "Smart

Thinking" Programme currently being piloted in the UK has also been specifically designed for use with young people and sets out to develop social and moral reasoning skills in young people as part of a group-work discussion process exploring issues surrounding offending behaviour. However, up until relatively recently there has been little in the way of practically applicable assessment tools for researchers seeking to evaluate the "moral reasoning component" of such programmes. Indeed, although the recent Home Office documentation outlining guidelines for evaluators of cognitive interventions provides a list of recommended psychometric tests it does not identify a specific tool for measuring moral reasoning (Colledge et al 1999:31-32). Part of the problem may stem from the fact that traditional measures of moral reasoning such as Kohlberg's Moral Judgement Interview (see e.g. Colby & Kohlberg, 1987) are impractical for use with large samples. Administration and assessment of the Moral Judgment Interview is complex, involving a lengthy structured interview process and requiring familiarisation with an intricate scoring procedure.

1.3 The Sociomoral Reflection Measure-Short Form

The SRM-SF that is the focus of the current study has several advantages over traditional measures of moral reasoning in terms of user-friendliness and has been designed for use with young people. The SRM-SF also has the advantage of being directly relevant to the assessment of moral reasoning enhancement within the context of a cognitive intervention with young people. The first author of the measure, John Gibbs, has provided the major contribution to the "sections" on moral reasoning in both the ART programme (Goldstein et al, 1998) and the EQUIP programme (Gibbs et al, 1995). Indeed, use of the SRM-SF as an assessment tool forms part-and-parcel of the EQUIP programme. However, as with much of the work in this field, the main body of empirical evidence concerning the use of the SRM-SF with young people derives from research that has been carried out in the US. Whilst there have been some recent UK-based studies (e.g. Palmer & Hollin, 1998) that have compared the moral reasoning of offending and non-offending adolescents using the SRM-SF, there is a need for further work in this area, particularly with reference to younger participants. Furthermore, the assessment of "moral reasoning" in young people brings with it a number of methodological and theoretical issues that, in the author's opinion, have not as yet been clarified by existing work in the

field. It therefore seems timely to carry out a comprehensive UK-based investigation of the moral reasoning of offending and non-offending adolescents using the SRM-SF.

1.4 Theoretical Background to the SRM-SF

The SRM-SF is a "production" measure of moral reasoning that has been developed by John Gibbs and colleagues (Gibbs et al., 1992). The measure has as its theoretical basis a revised version of Lawrence Kohlberg's stage model of moral development. Over the years, Kohlberg's theory of moral of moral development has aroused considerable controversy and debate. Today, nearly half a century after Kohlberg started his groundbreaking research, it is probably true to say that there are hardly any psychologists who would consider themselves to be Kohlbergians in the strictest sense. However, few researchers would fail to acknowledge the invaluable contribution that Kohlberg has made to the direction of research in this field. This section reviews key aspects of Kohlberg's theory and provides an overview of the revised Kohlbergian model put forward by Gibbs et al. (1992).

1.41 Overview of Kohlberg's Theory of Moral Development

Kohlberg's justice-oriented stage theory of moral judgement centres on developmental changes in moral reasoning. Kohlberg's model is a cognitive one, and follows Piaget in its underlying phenomenological, constructivist and structuralist assumptions. Moral judgements define and predict moral action. Before one can define an act as moral, one needs to know something of the reasons behind the act. If one is to understand and predict when and why people will "do right" one needs to gain insight into the way in which they think about "doing right". For Kohlberg, moral development occurs not merely as a result of the passive internalisation of cultural norms, but as an active constructive process. By thinking about and acting on the world, individuals construct moral meaning for themselves. Moral development is viewed as a process that takes place by progression through a sequence of stages characterised by certain underlying cognitive structures that are universal and invariant. Whilst social and cultural factors may influence the rate of progression through stages, and the content of moral judgements; the way in which moral judgements are structured within each stage and the sequence of development followed is perceived to be the same for all cultures. For Kohlberg, then, it is the form of thinking or the "general organising principles" surrounding beliefs and opinions

that exhibit developmental regularity and are generalised within and across individuals, not the individual beliefs and opinions themselves.

Moral development is seen as akin to cognitive development that occurs in logicomathematical and social perspective taking domains, with acquisition of the higher moral stages requiring higher levels of logical and social cognitive sophistication. However, high levels of logical and social cognitive sophistication are not sufficient in themselves for moral advancement. In order to achieve one's full moral potential one needs experiential exposure to role-taking opportunities in the social world. Role-taking opportunities occur through ones participation in various social groups, for example as a member of a peer group, member of a family and as a member of society. It is the level of social participation rather than the nature of the particular group that is seen as influential in the stimulation of moral development. For Kohlberg, parental socialisation is not critically necessary or unique for moral development as the necessary social stimulation may be provided by other sources (Kohlberg, 1984:75).

Kohlberg considered each moral stage to represent a structural whole, however, he also acknowledged that this presented challenges when seeking to account for mechanisms of transition from one stage to the next. Colby and Kohlberg (1987) noted that researchers in the field tend to distinguish between a transformational model of stage development and an additive or layer cake model. In the extreme form of a transformational model, the higher stage displaces the structure of the lower stage, and information in the lower stages is no longer accessible. Within the transformational model each developmental stage is perceived as strongly holistic, this implies great internal consistency in moral reasoning. In an additive or layer-cake model, higher stages are added to the repertoire with no loss of information from the lower stages. This model is associated with greater heterogeneity of stage use (Colby and Kohlberg, 1987:7).

Kohlberg when describing his position on the "Transformational model" versus "Additive model" of stage development claimed that the development of moral judgement as a whole was too broad in scope to be described by a single model. He argued that moral judgement includes a number of

distinct modes or processes including comprehension of moral judgements made by others, evaluation of judgements made by others and spontaneous production of moral judgements in response to questions about what is wrong and why (Colby and Kohlberg, 1987:7).

Kohlberg did not consider that the development of moral reasoning comprehension would conform to a transformational model – as it is clear that people at higher stages of moral development can comprehend moral judgements at lower stages, and "sometimes" may comprehend a stage or two above their own (Colby & Kohlberg, 1987:7). He also suggested that evaluation of or preference for moral judgements made by others might also fail to conform to a transformational model. This is because lower stage judgements "may often retain validity when seen from a higher stage perspective; the higher stage individual may attribute different meaning to lower stage judgements, but this does not prevent higher stage subject from endorsing as valid many ideas that are valid at lower stages" (Colby & Kohlberg, 1987:7). Kohlberg believed a transformational model was, however, appropriate for describing the "spontaneous production" of moral judgements.

Kohlberg qualified his position on the "Transformational" versus "Additive" model of stage development by making the distinction between moral competence and moral performance. He acknowledged that people do not always perform at their highest level of moral competence. However, he further argued that insofar as one was able to elicit moral competence through his "Moral Judgment Interview" technique (Kohlberg's moral judgment assessment tool), one would expect to find internal consistency in moral reasoning across a variety of moral dilemmas. He contrasted his transformational model of moral competence with additive models of social perspective taking. He noted that for some types of social perspective taking problems responses might vary across different levels even if it was the highest level of competence that was being assessed, as for some problems the lower level was all that is required to solve the problem. He argued that in the moral domain lower levels of reasoning were never morally more adequate, even if under certain circumstances they were psychologically more adequate. He argued that individuals would only use lower than optimum levels of moral judgement in situations of significant downward moral press (for example, offenders living in the "low-level" moral atmosphere of a prison may exhibit lower levels of moral reasoning on hypothetical prison dilemmas than on other moral dilemmas). However,

Kohlberg did acknowledge towards the end of his life that performance variables influential in the expression of moral judgement had only just begun to be delineated and that was an area that needed further work (Colby & Kohlberg, 1987:8).

1.42 Description of Kohlberg's Moral Stage Typology

Kohlberg's moral typology consisted of three levels, each level incorporating two developmental stages: The "preconventional" level (encompassing moral stages 1 and 2), the "conventional" level (encompassing moral stages 3 and 4) and the "postconventional" level (encompassing moral stages 5 and 6). Kohlberg suggested that the differences between the three levels might best be understood by interpreting them as representing a changing relationship between the self and society:

> *"Level 1 (preconventional) is a perspective from which rules and social expectations are something external to the self; in the Level 2 perspective the self is identified with or has internalized the rules and expectations of others, especially those of authorities. and the Level 3 (postconventional) perspective differentiates the self from the rules and expectations of others and defines moral values in terms of self chosen principles".*

(Colby, Kohlberg et al, 1987:16)

1.42i The Preconventional Level

Kohlberg viewed the preconventional level as typifying the moral reasoning of most children under 9, some adolescents and many adolescent and adult offenders (Kohlberg, 1976). At this level, whilst the child is aware of and is able to respond to cultural concepts of "right and wrong", moral values are perceived in terms of the physical consequences of an action or with reference to instrumental concerns. At Stage 1, (the "Punishment and Obedience" Orientation) the "goodness or badness" of a moral act is defined by rewards or punishments associated with the act or with reference to the status of the actor. At Stage 2, (The "Instrumental Relativist" Orientation) moral values are interpreted in terms of direct exchanges, pragmatic needs and hedonistic concerns. Frequently, moral justifications are presented as "cost/benefit" analyses weighing up potential advantages and disadvantages for the individual. As Kohlberg put it, "human relations are viewed in terms like those of the market place"

(Kohlberg, 1981: 17). At the preconventional level, then, moral awareness is "external to the self" in that moral values tend to be conceptualised according to direct outcomes of a moral act or with reference to superficial attributes of the actor.

1.42ii The Conventional Level

Kohlberg's "conventional" level is deemed to be characteristic of the type of moral reasoning attained by most adults in most societies. At this level, the individual has come to "internalise" moral values, in the sense that maintaining the social expectations of others are now valued in their own right, regardless of the "immediate and obvious outcomes" (Kohlberg, 1981:18). At Stage 3 ("Interpersonal Concordance" or "Good Boy / Nice Girl" Orientation) moral acts are perceived as those that help, please and are approved of by other people. It is at this stage that the intentions of the actor become relevant when defining an act as moral. As Kohlberg puts it "the judgement 'he means well' becomes important for the first time" (Kohlberg, 1981, 18). At Stage 4, (Society Maintaining Orientation) interpersonal considerations are extended to encompass society's expectations. Moral acts are interpreted with reference to ones responsibility as a member of society and the individual who reasons at this stage has internalised society's norms. At this stage, "Doing ones duty as a citizen", "maintaining social order" and "showing respect for legitimate authority" are valued in their own right (Kohlberg, 1981).

1.42iii The Postconventional Level

Kohlberg's postulation of a "post conventional" level of moral reasoning is one of the most contentious aspects of his theory. The postconventional level represents an "ideal" in terms of moral maturity, and has been identified in only a minority of adult individuals in certain "complex" societies. The "post conventional" reasoner has moved beyond the need to define moral values according to interpersonal or societal norms. At this level, moral values become conceptualised as principles that are valid in their own right. These principles are independent from the authority of the people holding these values and from the relationships that the principled reasoner has with people holding these values. At Stage 5 (The Social Contract Orientation) "morality" is conceptualised in terms of individual rights and / or in terms of commonly held standards that have been arrived upon

through critical debate. The Stage 5 reasoner acknowledges the relativism of personal values, is likely to question commonly held standards and emphasises the possibility for changing law, when necessary, as part of a democratic process. Kohlberg claimed that "Stage 5" was the official morality of the American Government and Constitution (Kohlberg, 1981: 19). At Stage 6 (The Universal Ethical Principle Orientation) – the highest stage in Kohlberg's typology – morality becomes embedded in the self. "Right actions" are those that conform to personal conscience and that are in accord with an individual's self-constructed moral principles. These principles are ones that "appeal to logical comprehensiveness, universality and consistency". They are "abstract" and "ethical" rather than a set of concrete rules such as the Ten Commandments. Kohlberg saw these principles as

> *"... universal principles of justice, of the reciprocity and equality of human rights and of respect for the dignity of human beings as individuals"*
>
> *(Kohlberg, 1981, 19)*

1.43 Kohlberg's Moral Judgement Interviews: The "Heinz Dilemma"

In order to explore individual differences in moral reasoning, Kohlberg constructed a series of moral dilemmas- the best known of these being the so-called "Heinz dilemma". There are some variations to the Heinz dilemma, but broadly speaking the scenario is as follows:

"Heinz's wife is dying from a serious disease. There is a drug that the doctors think can save her, a drug that has been developed by a pharmacist living in the same town as Heinz. However, whilst the drug is expensive to make, the pharmacist is charging ten times what it cost him to make. Heinz goes to everyone he knows to try and borrow the money for the drug, but can only raise half the amount he needs. Heinz goes to the pharmacist and explains his situation, asking the pharmacist to reduce the price. The pharmacist refuses. Heinz gets desperate and breaks into the pharmacy to steal the drug."

Participants in moral judgement interviews are first asked whether Heinz should have stolen the drug and then a series of probing questions in order to elicit moral justifications – for example: "Why should/shouldn't Heinz have stolen the drug?" "Should Heinz be punished for what he did?" "Why should/shouldn't Heinz be punished for what he did?" Kohlberg was interested in the justifications

that participants gave for the decisions they arrived at rather than the actual decision itself. Responses to the Heinz dilemma may be coded according to Kohlberg's moral stage typology. Someone reasoning at Stage 1 ("Punishment and Obedience Orientation") might justify the decision for Heinz to be punished for stealing the drug by saying: *"He'd be a thief if they caught him and the police would put him in jail"* (Kohlberg, 1981: 257), punishment being seen as an automatic response to a bad act. However, someone reasoning at Stage 5 (The Social Contract Orientation) might justify not punishing Heinz by making reference to the need for the judge to take into consideration the special circumstances surrounding this case. Or in Kohlberg's words: *"The judge's duty is not merely to apply general rules to particular cases but to do so in such a way that fairness to individual disputants is achieved and the underlying purposes or polices of the law are served"* (Kohlberg, 1981: 236).

1.44 Empirical Investigations of Kohlberg's Theory: The Longitudinal Study of US Males

Kohlberg and colleagues carried out several empirical studies to test his stage model of moral development (see e.g. Colby, Kohlberg et al, 1987 for reviews of the early work). Possibly the best known of these studies is Kohlberg's longitudinal investigation of age trends in moral reasoning among US males. This study – which was started in 1956 - followed the progression of 50 males (aged 10, 13 and 16 at outset) over a period of 20 years. Participants' moral reasoning was assessed every 3 years using Kohlbergian moral dilemmas. The final outcome of this project provided some evidence to support claims that moral reasoning developed according to Kohlberg's typology. In general, over time, it was found that individuals' "preconventional" reasoning diminished with age, whereas "conventional" reasoning increased with age. Additional support for the notion that moral development follows an invariant sequence came from further studies that demonstrated that advances in moral reasoning were almost always to the next stage up (reviewed in Colby, Kohlberg et al 1987), with little evidence to suggest that individual skipped stages. However, it should be noted that during the course of the study, Kohlberg had to make refinements both to his typology and scoring scheme in order to address issues that arose at intermediate stages of the project. For example, using his original scheme, Kohlberg found that a substantial minority of his participants demonstrated apparent "Stage regression" at an intermediary follow-up period. In particular, some individuals who had been scored

at the post-conventional level (Stage 5 or Stage 6) during high school were later scored as being at the preconventional level Stage 2 when retested as college students. This observation was clearly contra to the theory that the development of moral competence corresponded to a "transformational" stage model. Kohlberg and colleagues attempted to reconcile this apparent conflict with reference to certain complexities in coding moral protocols. Redefining the coding scheme to produce a standardized Structural Issue Scoring scheme resolved many of the methodological difficulties, yielding "more orderly" data (Colby & Kohlberg, 1987:39). However, in the course of resolving methodological problems associated with scoring moral protocols, Kohlberg may have created a new theoretical problem relating to the distinction between the "conventional" and the "postconventional" levels (Gibbs et al, 1992:15-16). This issue will be discussed in more detail when considering the Gibbs et al (1992) revisionist model of moral reasoning later in this chapter.

1.45 Empirical Investigations of Kohlberg's Theory: Cross-Cultural Studies

Kohlberg and others have carried out a number of studies in a variety of different countries in order to test the claim that his typology represents a universal culturally invariant model of moral development. The majority of work that has been carried out has involved cross-sectional comparisons of age differences in moral reasoning across cultures; however, some longitudinal studies have also been carried out in different cultures. These include a longitudinal study carried out amongst Turkish males (Nisan & Kohlberg, 1982) and a longitudinal study of kibbutz adolescents in Israel (Snarey, et al, 1985a). The evidence that has emerged from the cross-cultural studies has provided qualified support for some of Kohlberg's underlying claims. Snarey (1985) in his review of 45 different studies in 27 different countries concluded that the evidence pointed to a certain universality for Kohlberg's scheme when considering his preconventional and conventional levels of moral reasoning. However, Snarey found little evidence to support the claim for universality of Kohlberg's postconventional level of moral reasoning. Cross-cultural evidence of this level was rare, and when it was found was restricted to middle-class populations in complex urban societies.

Whilst Kohlberg's approach to moral reasoning emphasised "naturalistic" developmental processes he also postulated that optimal moral development could only be achieved in a socially rich environment

– the greater the level of stimulation through social participation the greater the rate of moral development. For Kohlberg, potentially important markers of social role-taking opportunities included socio-economic status of origin, sociometric status and level of formal educational attained. In his longitudinal study of US males, Kohlberg found moderate to strong effects for social class and formal educational level as mediators of moral reasoning level (though only very limited effects for sociometric status). Kohlberg interpreted differences in socio-economic status and formal educational level as representing "differential participation in and identification with the society and its secondary institutions" and argued that these differing levels of participation create differing role-taking opportunities (Colby & Kohlberg, 1987:113). Some researchers have sought to relate the failure to evidence postconventional reasoning cross-culturally to the "social participation model"; however, as Snarey noted, seeking to rationalise the findings in this way raised the issue of cultural ethnocentrism in the Kohlbergian model (Snarey, 1985:227). As Snarey pointed out, there is no reason to suppose that members of so-called "tribal" or folk communities do not experience the social prerequisites for the development of postconventional moral reasoning. Snarey argued that a theory of this nature even in its mildest form would apparently support a misguided doctrine of "social evolution" with so-called "primitive" societies at lower levels of the hierarchy and complex urban societies such as the US at the top (Snarey, 1985). Snarey drew attention to the fact that whilst Kohlberg's description of the preconventional and levels of moral reasoning were well based on "empirical operative judgments", his description of the postconventional level was primarily drawn from the writings of Western philosophers. Snarey concluded on the basis of his research review that whilst there was considerable evidence to support Kohlberg's underlying assumptions relating to the preconventional and conventional moral levels, his description of the postconventional level was probably incomplete and required supplementation from a wider range of cultural world views (Snarey, 1985:228-229).

1.46 The Relationship between Moral Reasoning and Offending Behaviour

A key assumption underpinning Kohlberg's theory is that moral reasoning is an important determinant of moral action. However, Kohlberg did not himself elaborate on this assumption by providing a formal definition of the relationship between moral judgment and moral conduct. In acknowledging that people do not always do what they think to be right, he merely noted that the relationship between moral judgment and moral action is "complex and incompletely understood" (Colby & Kohlberg, 1987:2). Whilst Kohlberg stressed the point that moral reasoning defined and predicted moral behaviour, it is also clear from his theory that he did not expect there to be a "perfect" relationship between reasoning and action. If one considers potential responses to the Heinz dilemma, one finds that individuals reasoning at disparate developmental levels may favour the same course of action (for example "stealing the drug"), whilst justifying the course of action in different ways. Conversely, individuals reasoning at the same level of moral development may favour different courses of action using structurally similar types of moral justifications.

Blasi (1980) pointed out that cognitive-developmental theory was vague when it came to approaching the relationship between cognition and action (Blasi, 1980:1). Blasi set out to address this issue by reviewing the available empirical research in the field. He concluded that whilst there was considerable support for the hypothesis that moral reasoning and moral action were related, the level of support varied according to the domain under investigation (Blasi, 1980:37). The empirical studies that provided most support for the claim that moral reasoning related to moral action came from the work that investigated between-group differences in the moral reasoning of delinquents and non-delinquents. In a review of 15 studies comparing delinquent and non-delinquent samples, Blasi found that in 10 of these there were significant differences between the two groups with delinquents demonstrating lower levels of moral reasoning than non-delinquents (Blasi, 1980:12).

Blasi noted that Kohlberg located differences in the moral reasoning between delinquent and non-delinquent adolescents in disparities in the use of preconventional and conventional levels of moral reasoning. Kohlberg believed that the majority of adolescent offenders were functioning at the preconventional level of moral reasoning, whereas the majority of "average" adolescents and adults

who understood, accepted and attempted to uphold the values and rules of society were functioning at the conventional level of moral development (Kohlberg et al., 1975:243). However, when Blasi explored this proposition within the context of his own review his findings were equivocal. Whilst some studies he reviewed indicated that the majority of offenders were functioning predominantly at the preconventional level there were other studies reporting that a substantial number of adolescents scored at Kohlberg's conventional level of moral reasoning (Blasi 1980:12). Jurkovic, carrying out a similar review study in 1980 concluded that empirical support for the claim that delinquents reasoned "immaturely" about moral issues varied considerably both across and within studies (Jurkovic, 1980:719). Furthermore, Jurkovic pointed out that findings from some studies that investigated the moral reasoning of delinquents indicated that there was evidence of considerable inter-individual and intra-individual variation in moral reasoning – especially amongst those who appeared capable of conventional moral reasoning (Jurkovic, 1980:717).

The studies carried out by Blasi and Jurkovic in 1980 were, essentially, narrative reviews of the available empirical research. In later work using more sophisticated statistical techniques of analysis greater credence was given to the Kohlbergian view that offending and non-offending adolescents could be distinguished in terms of preconventional and conventional levels of moral reasoning. There is now a substantial body of research that has established that in terms of moral reasoning offenders are "developmentally delayed" in comparison with their non-offending peers (see e.g. the meta analysis carried out by Nelson et al, 1990; also studies carried out by Lee & Prentice,1988; and Chandler & Moran, 1990). However, it is perhaps worth flagging the equivocal findings of Blasi (1980) and Jurkovic (1980), as one tends to find that these early reviews are often cited in the research literature as providing strong evidence for Kohlberg's claims.

1.47 Moral Reasoning Interventions with Young Offenders.

As was noted in the beginning of this chapter, moral reasoning enhancement forms one of the core components of many of the most widely used multi-modal cognitive interventions currently used with offenders in the UK. However Kohlbergian moral reasoning enhancement has been used as an intervention with young offenders in its own right almost since the inception of Kohlberg's theory of

moral development. Interventions designed to raise levels of moral reasoning in young offenders fall into two categories: microinterventions and macrointerventions (Hollin et al, 2002:39). Microinterventions typically focus on seeking to enhance the moral reasoning of participants as part of a group-work process. For example, the "Smart Thinking" programme that was being piloted at the time of the current study presents young people with a number of differing scenarios involving moral dilemmas. Facilitators encourage moral reasoning enhancement through a guided discussion process that presents participants with challenges to existing moral beliefs.

Macrointerventions are more ambitious than microinterventions in that they seek not only to enhance the moral reasoning of participants, but also to enhance group members' "moral climate". It will be recalled that Kohlberg believed that in order for optimal moral development to occur, individuals needed to have access to a socially rich environment. In some early studies investigating the impact of moral reasoning interventions in prisons, Kohlberg noted that whilst there was evidence of some positive outcomes in terms of moral reasoning enhancement, overall the effects were weak (e.g. Kohlberg et al 1975:244). Kohlberg was, however, not surprised by this, bearing in mind the "moral environment" of traditional prisons. He considered that many prisons operated at a "significant downward moral press" (Colby & Kohlberg, 1987:8, see also Scharf & Hickey, 1976). Kohlberg and others encouraged the setting up of "democratic" or "Just Community" units within the prison environment, where prisoners were encouraged to actively participate in and contribute to certain aspects of the running of the unit. The "Just Community" approach has also been adopted as a means of enhancing the moral climate of some schools (see e.g. Kuther & Higgins-D'Alessandro, 2000).

As Hollin et al (2000) point out, whilst there have been a number of studies that have demonstrated that group work interventions can enhance the moral reasoning of offenders, there have been relatively few studies that have examined the impact of moral reasoning enhancement per se with reference to behavioural outcomes. However, in a study carried out in 1986, Arbuthnot and Gordon investigated the effects of a cognitive moral reasoning based programme on the behaviour of American adolescents previously identified as at "high-risk for delinquency". Individuals participated in a 16-20 week programme of moral reasoning education that included discussion sessions and role

playing opportunities. Arbuthnot and Gordon administered a variety of pre and post tests to both the experimental and the control group. Whereas there were no significant differences between control and experimental group on any of the measures at pre test, at post test significant differences were observed between the two groups on a number of the measures. There was a significant improvement in moral reasoning in the experimental group. There was also a significant reduction in behaviour problems among the experimental group, with fewer referrals to the school office for behaviour problems, fewer incidences of lateness, and far fewer reported contacts with police or juvenile courts (Arbuthnot & Gordon 1986).

1.48 Potential Gender Bias in Kohlberg's Typology

One of the major criticisms of Kohlbergian theory that has in itself aroused considerable debate and controversy is that the theory is gender-biased and represents a male oriented view of morality. The best known and most prominent figure to initially make this claim was Carol Gilligan (see Gilligan, 1982, 1998). Gilligan claimed that Kohlberg's justice oriented model of morality failed to acknowledge the importance of the "ethic of care" prevalent in women's dialogues about moral issues. Whilst Gilligan's theory is perhaps most commonly associated as representing a feminist perspective on psychology, it is interesting to note that her initial concerns regarding the neglect of the "care ethic" in Kohlberg's theory arose from her observations of young <u>males</u> confronted with real life dilemmas. Her first intimation that there might be a problem with Kohlberg's theory of moral development came when she was co-teaching a course of moral development with Kohlberg during the Vietnam era. Gilligan noticed that the young men in her class were reluctant to talk about the draft, apparently aware that there was no room in Kohlberg's theory to talk about their feelings without sounding morally undeveloped (Gilligan 1998:128).

In Kohlberg's typology one finds that at his conventional level a developmental progression is made between Stage 3 ("Interpersonal Concordance" or "Good Boy / Nice Girl" Orientation) and Stage 4 (Society Maintaining Orientation). Gilligan argued that the differences between Stage 3 and Stage 4 represented qualitative differences in styles or voices of morality rather than different developmental stages of moral reasoning (Gilligan 1982). For Gilligan, Stage 3 approximated to a "Care and

responsibility" ethic characteristic of the "female voice", and Stage 4 corresponded to a "Justice and rights" ethic more characteristic of males. Gilligan attributed gender differences in moral voices to differential patterns of upbringing experienced by males and females. Gilligan argued that males are socialized to become independent and assertive whereas females are encouraged to become empathic and nurturing (Gilligan 1982).

The implication of Gilligan's hypothesis, then, is that when assessing differences in conventional styles of moral reasoning adult females will typically score at a lower moral stage (Stage 3) than adult males (Stage 4). A review of the literature reveals that evidence for Gilligan's claims is conflicting. Certainly in some early studies there was evidence to indicate that females tended to do less well than males on Kohlberg's Moral Judgment Interview, and that in some cases this could be attributed to greater evidence of Stage 3 reasoning amongst females (see e.g. Bussey & Maughan, 1982 and Ford and Lowery, 1986 for reviews). However, in an influential meta-analysis published in 1984 Walker concluded that if one controlled for occupation and educational differences, there were no "non-trivial" gender differences in use of Stage 3 and Stage 4 reasoning (Walker, 1984; but see also Baumrind, 1986 for a rejoinder to Walker, and rebuttal by Walker, 1986). Jaffee and Hyde in a meta-analysis of gender differences in moral reasoning (Jaffee & Hyde, 2000) concluded that evidence to support Gilligan's claims was at best limited. However, they did note that gender effect varied according to age, type of moral reasoning measure used and method of analysis. More recently Tangney and Dearing have observed that when gender differences do occur in moral reasoning, they invariably correspond to the Gilligan hypothesis. They report that they know of no studies where adult males demonstrate greater evidence of the care ethic than adult females (Tangney and Dearing 2002:264).

Garmon et al (1996) claim that an important distinction to make in the research on gender differences in moral judgement is between "stage level or structure, and expressive emphasis or content" (Garmon et al, 1996:418). It will be recalled from the overview to Kohlberg's moral judgement typology that for Kohlberg internal consistency in moral judgement related to the way in which moral judgements were structured, not in the individual content of these moral judgements. Along with previous researchers, (e.g. Walker et al 1987) Garmon et al (1996) argue that Gilligan's critique of

Kohlberg's typology may be addressed as two separate research questions. Firstly, the question of whether the typology is stage biased against females. Secondly, the question of whether there are gender differences in moral orientation (or content) within each global moral stage (Garmon et al, 1996:419). Garmon et al (1996) note that whilst initial research investigating gender differences in moral judgement focussed on investigating differences in structural or stage level, more recently attention has turned to the investigation of differences in content of moral judgement within each stage level. Drawing from the previous research literature and on the basis of their own empirical work, Garmon et al (1996) discount Gilligan's claim that Kohlbergian measures of moral reasoning are gender biased in structural terms (i.e. in terms of Stage 3 versus Stage 4 reasoning); however, they argue that there is partial support for the claim that there are gender differences in the qualitative content of moral judgements within each global moral stage. In their investigation of gender differences in stage and expression of moral judgement using SRM-SF data Garmon et al (1996) found no evidence of Stage bias favouring males. Indeed, they found evidence to indicate that during early adolescence that females were out-performing males on the measure. However, they did find evidence to indicate that within-Stage judgments varied in orientation according to gender, the content of moral judgements of females being more care focussed than those of males.

Clearly, for those subscribing to a strong version of the Gilligan (1982) hypothesis, the apparent distinction between structure and content of moral judgements made by stage theorists may not represent a straightforward distinction to make. For Gilligan, Stage 3 is by its very nature more care focussed than Stage 4, which in turn is deemed to have a strong "justice" emphasis". As noted earlier, for Gilligan, differences in Stage 3 and Stage 4 reasoning in themselves represent qualitative differences in styles of reasoning that should not be conceptualised as different points on a hierarchical typology. Rather, these differences are better represented as parallel pathways of moral development (Gilligan 1982:18-23). Nevertheless, in the current study, in line with Garmon et al (1996) the distinction is made between structure and content of moral judgements. The empirical investigation of gender differences that forms part of this thesis is restricted to the investigation of structural differences in global moral stage use, not to the investigation of gender differences in moral content or orientation within stages. The rationale for this approach is two-fold: firstly, the bulk of the

research on gender differences in moral stage use has been conducted amongst adults, less attention has been paid in the literature towards potential gender differences in moral stage use amongst adolescents; secondly, in the author's opinion, there may be some theoretical issues regarding the coding of SRM-SF data by content category within stages. Some of the potential theoretical issues relating to the content categories of the SRM-SF global stage typology (so-called "Aspects") are taken up in the Discussion in Chapter 3 and are further elaborated upon in the Global Discussion that forms the final chapter of the current thesis.

1.49 The Relationship Between Gender, Offending Behaviour and Moral Reasoning

As is well documented in the offending behaviour literature, antisocial behaviour in young people is much more prevalent amongst young males than it is amongst young females (see Rutter et al, 1998, for a review). However, on the face of it, it would seem that the moral reasoning research offers little in the way of providing an explanation for these differences. If one bears in mind the theoretical link between moral reasoning and offending behaviour, the apparent tendency for females to do less well than males on moral reasoning tasks would seem to be contradictory. As was noted earlier, Kohlberg's theory of moral development has been challenged on the grounds that it is gender-biased and that it has little to say about female morality (e.g. Gilligan 1982, 1998). Gilligan and Wiggins (1987) suggest that failure to identify gender differences favouring females in the moral reasoning domain may relate to issues of measurement. Typically, in the moral reasoning research one finds that findings are presented with reference to indices such as summary mean scores. It will be recalled that in some studies females have evidenced greater use of Stage 3 reasoning than males, and that males have evidenced greater use of Stage 4 reasoning than females. If it is the greater use of Stage 3 amongst females that is key to understanding gender differences in antisocial behaviour, then one may not be able to evidence this through between-group comparisons of summary mean scores. Potential measurement issues relating to the SRM-SF are taken up later in this chapter.

1.5 Gibbs' Revision of Kohlberg's Theory

The moral reasoning measure used in the current study (the SRM-SF) is derived from John Gibbs' revisionist model of Kohlberg's theory of moral development (see e.g. Gibbs, 1977; Gibbs & Schnell,

1985, Gibbs et al, 1992). As with Snarey (1985), Gibbs highlights the difficulties associated with Kohlberg's postconventional level of moral reasoning; however he comes to a different theoretical conclusion to Snarey. For Gibbs, the postconventional level is problematic not so much because it represents the writings of a Western philosophical elite, but that it appears to be an existential or reflective extension of the earlier stages rather than representing a naturalistic developmental progression (Gibbs, 1977). For Gibbs, then, supplementing the postconventional level with principles from non Western societies would not be enough to resolve the problems with this level as it would still confuse moral maturity with philosophical articulateness (Gibbs et al, 1992:19). Furthermore, Gibbs et al (1992) note that when redefining his typology as a result of anomalies observed in the intermediate phase of his longitudinal study, Kohlberg created a potential contradiction in the theoretical definitions of the conventional and postconventional levels. In order to resolve some judgments initially coded at the postconventional level were subsequently recoded as being at the conventional level at a follow-up period. These included justifications such as: "the moral value of life takes precedence over obedience to laws or authority". By doing this, however, in Gibbs' view he created a conflict between the inclusion of justifications that apparently corresponded to "principled-sounding moral ideality" within Stages 3 and 4 and the traditional definition of these stages as conventional (Gibbs et al, 1992: 15-16). Gibbs et al (1992) do not believe that Kohlberg resolved this issue. In their view, both Stage 3 and Stage 4 already demonstrate mature moral development. They have re-cast Kohlberg's original 3 level, 6 stage classification into a 4-stage system: Stages 1 and 2 represent "immature" levels of moral reasoning, Stages 3 and 4 represent "mature" levels of moral reasoning. Progression between the major stages is conceptualised by the incorporation of "transitional" levels between Stages 1 and 2, Stages 2 and 3 and Stages 3 and 4. Further delineations between moral justifications are provided within each developmental stage by the inclusion of content categories known as "Aspects".

Like Kohlberg's MJI, the SRM-SF is a production measure of moral reasoning. Production measures differ from recognition measures in that production measures are designed to elicit spontaneous production of moral judgements whereas recognition measures require individuals to evaluate or rank examples of moral justifications. This difference is important when one considers Kohlberg's

consideration of the "Transformational" model versus the "Additive" model of stage development and his distinction between moral competence and moral performance. It will be recalled that Kohlberg considered the Transformational model to apply to the spontaneous production of moral judgements – or moral competence – but not to the evaluation or comprehension of moral judgements. The SRM-SF is, then, designed to tap into moral competence, and one would predict from this that individuals would be relatively internally consistent in their patterns of moral stage use.

The SRM-SF differs from the MJI in that it does not make use of hypothetical dilemmas in order to elicit moral judgments. Instead, participants are first asked to evaluate the importance of, say, "Keeping a promise to a friend" and then to give reasons to justify their evaluation.
As Gibbs et al (1992) note, over the years, the use of moral dilemmas as a means of assessing moral reasoning has been criticised by a number of researchers as being "artificial, inappropriate, or irrelevant for children" (Gibbs et al, 1992:37). The straightforward format of questionnaire items in the SRM-SF makes it more suitable for use with younger participants than other production measures of moral reasoning. Pilot studies of the SRM-SF indicated that it may be used with children as young as 8 years old . From a practical point of view, the SRM-SF is easier to administer and analyse than other production measures, requiring only about 30 hours of initial training.

1.6 Description of Moral Stages in the SRM-SF

Broadly speaking, progression through stages is seen as a cognitive process of increasing "decentration" from the superficial to the profound. Stage 1 is characterised by the young child's emphasis on salient, concrete, physical and inevitable aspects of moral events. By Stage 2, the child is able to incorporate psychological perspectives in their interpretation of events. However, these psychological perspectives are "superficial" in that the child focuses primarily on pragmatic, egoistic and instrumental concerns. Moral maturity is achieved once the person's interpretation moves beyond the superficial in order to be able to infer the bases of interpersonal relationships (Stage 3) or societal norms (Stage 4). This section provides a description of the SRM-SF stages as presented in Gibbs et al (1992: 21-31).

Stage 1 - Unilateral and Physicalistic:

At this Stage, sociomoral justifications are seen as being "extrinsic" and "authority-oriented". Typically, the explanations that the young person gives to justify moral actions will make reference to a powerful authority figure, will be expressed in terms of absolutes and will often allude to the inevitable physical consequences of the act. The emotional consequences of moral behaviours, insofar as they are recognised, are described in terms of their overt expression, rather than with reference to the underlying feelings of individuals.

Transition Stage 1/2

The change from Stage 1 to Transition 1 / 2 reasoning is characterised primarily by a shift from the inevitability of physical outcomes to a more probabilistic outlook. As Gibbs and colleagues put it: "External consequences become physical or inevitable, but not both." Linguistically, this shift tends to be manifested in a change from the use of the simple future – e.g. "you will get beaten up" - to the conditional – "you might get beaten up".

Stage 2 - Exchanging and Instrumental

By Stage 2, the young person's perspective has moved beyond the immediately obvious, physical and salient aspects of events. In contrast to the unilateral authority-oriented reasoning of Stage 1, Stage 2 reasoning is characterised by an "autonomous morality of social interaction" (Gibbs et al, 1992:23). However, sociomoral concerns are restricted and immature in that they tend to be conceptualised in terms of "direct exchanges" and "pragmatic needs". Frequently, sociomoral justifications are presented as "cost/benefit" analyses weighing up potential advantages and disadvantages for the individual. Typically, reasoning at this stage is instrumental and calculative in style, with a strong emphasis on hedonistic concerns.

Transition 2/3

The change from Stage 2 to Transition 2/3 is perceived as a move away from the view of reciprocity as "fact" and a shift towards more hypothetical and intrinsic aspects of sociomoral behaviour. As with the change from Stage 1 to Transition 1 / 2, this shift manifests itself linguistically. Whereas

stage 2 exchange-type justifications are construed using the simple future tense – "they **will** return the favour", Transition 2/3 makes more use of the future subjunctive - "you **would want** them to keep a promise to you". It is during this stage that one first sees reference to intrapersonal considerations such as "conscience" being offered in justifications for moral behaviour. However, at this stage "conscience" is not formally internalised, being conceptualised instead as an external, potentially bothersome agent that would, for example, "hound" the individual.

Stage 3: Mutual and Prosocial

By Stage 3, the mature moral reasoner has ceased to focus on the superficial and the external aspects of moral behaviour. Social relationships are no longer perceived as types of business exchanges but instead are construed in terms of their intrinsic nature and worth. By this Stage, intrapersonal considerations such as conscience have been truly internalised. Justifications at this Stage reflect a deep understanding and consideration of the psychological and emotional welfare of others. However, moral justifications that move beyond the confines of specific relationships are also in evidence at this stage. Example of more "global" justifications are those that "generalise" care from individuals to the human race as a whole, or justifications that make reference to normative codes of conduct.

Transition 3/4:

As with the previous transitional phases, Transition 3 /4 represents a shift from one moral stage to the next. As such, transitional phases are characterised by reasoning styles that are "blends" of adjacent stages. For example, transition 3 / 4 is typified by judgments that are generalised beyond specific relationships but do not yet extend to the functional needs of a complex social system. However, Transition 3 / 4 is different from the other transitional phases of moral reasoning in that there is a distinct style of reasoning associated with this stage. This style of reasoning is referred to as "Relativism of Personal Values" (RPV). Reasoning that is characteristic of this aspect relates to personally held values and beliefs. For example, obeying the law might be seen as being important, but "one should act according to one's values". Gibbs and colleagues believe that many individuals who emphasise this aspect of reasoning to extremes may "eventually come to realize its anarchic

implications" and will proceed to construct Stage 4. However, they believe that for some people RPV may remain a relatively stable aspect over considerable periods of time.

Stage 4: Systemic and Standard

The maturity evident at Stage 3 is extended to address the functional aspects of complex social systems. Typically, judgements at this stage are characterised with reference to society's needs, in terms of social justice or with reference to moral standards such as "integrity" and "self-respect".

1.7 Reliability and Validity of the SRM-SF

In an extensive investigation of the psychometric properties of the SRM-SF Basinger et al (1995) tested the reliability and validity of the measure using a number of methods. They reported good levels of test / re-test correlations, internal consistency and inter-rater reliability of the SRM-SF. The questionnaire was also deemed to have achieved acceptable concurrent validity as Basinger et al (1995) found that the correlation between scores on the SRM-SF and the Moral Judgement Interview was highly significant. They also found the SRM-SF to achieve good construct validity in that the measure successfully discriminated between diverse age groups from 10 year olds to adults - indicating that the measure was sufficiently sensitive for exploring developmental age trends. Basinger et al (1995) also found that the measure was successful in discriminating between delinquent and non-delinquent adolescents. A number of other studies carried out in the US, have, broadly speaking, confirmed the findings that the SRM-SF achieves good levels of inter-rater reliability (e.g. Garmon et al 1996), at least moderate levels of internal consistency (e.g. Krcmar & Valkenburg, 1999), that the SRM-SF is sensitive to age (e.g. Humphries et al, 2000) and that the measure successfully discriminates between "delinquent" and "non-delinquent" adolescents (e.g. Gregg et al 1994). Finally, in three reported studies, Gibbs and colleagues found that there was no correlation between SRM-SF scores and measures of social desirability (Gibbs et al 1992; Gregg et al, 1994; Basinger, et al 1995). This suggests that the SRM-SF may be relatively resilient in terms of "faking bias" (c.f. Aleixo & Hollin, 1996).

As a measure, the SRM-SF's cross-cultural transferability has yet to be conclusively established. However, over the last few years findings have started to emerge from studies using non-US based samples that indicate that the SRM-SF's applicability is not restricted to a US cultural setting. Studies carried out by Comunian & Gielen in Italy found that an Italian version of the SRM-SF demonstrated acceptable levels of internal consistency and inter-rater reliability; Comunian & Gielen also found that the Italian version successfully discriminated between older and younger adolescents (Comunian & Gielen, 1995; Comunian & Gielen, 2000). Work carried out by Krettenauer and Becker indicated that a shortened German version of the SRM-SF was suitable for use with German adolescents. Krettenauer and Becker reported acceptable levels of internal reliability for their version of the SRM-SF and that scores on the measure correlated significantly with age; Krettenauer and Becker also found a significant relationship between scores on their measure and greater or lesser admissions of offending behaviour on a Self Reported Delinquency Scale (Krettenauer & Becker, 2001). Ferguson et al (2001) concluded from their study comparing SRM-SF scores of Nigerian and Northern Irish children that the SRM-SF was suitable for use in a Nigerian setting (Ferguson et al, 2001). Mizuno (1999) reports that the SRM-SF was suitable for comparing cultural differences in moral reasoning between Japanese and American adolescents. Finally, in Australia, the SRM-SF was found to be useful as a assessment tool evaluating the impact of a Victim Awareness Program with young offenders (Putnins, 1997)

1.71 Evidence of Reliability and Validity of the SRM-SF from UK Based Samples:

Research findings using UK based samples have also been published that indicate that the SRM-SF is suitable for use in a British cultural setting. An investigation of the SRM-SF's reliability and validity with reference to UK participants was carried out by Ferguson et al (1994) amongst a sample of Northern Irish children and young people. They reported high levels of test / re-test correlations and acceptable levels of internal consistency for the SRM-SF. They also report acceptable levels of convergent and concurrent validity for the SRM-SF. Ferguson et al (1994) found significant correlations between scores on the SRM-SF and scores on the Sociomoral Reflection Objective Measure and also that SRM-SF scores correlated with age. Further work carried out in Northern Ireland also found that the SRM-SF was apparently sensitive to age: Ferguson & Cairns (1996) found

significant differences between the SRM-SF scores of 10-11 year olds and those of 14 – 15 year olds. In mainland Britain, the bulk of the evidence providing support for the reliability and validity of the SRM-SF comes from research carried out by Emma Palmer and Clive Hollin which has explored dimensions of offending behaviour, parenting styles and moral reasoning amongst young people in England. Palmer and Hollin have found the SRM-SF to be effective at discriminating between young offender and non-offender samples (Palmer & Hollin 1998, Palmer & Hollin 1999). They have also reported significant relationships between SRM-SF scores and age amongst non-offending samples (Palmer & Hollin 2000, Palmer & Hollin 2001).

1.72 Gender Differences Identified by the SRM-SF

There is little evidence to suggest that the SRM-SF is susceptible to gender bias favouring males in terms of Moral Stage development scores. Indeed, if anything, the reverse is the case - as in general, where gender differences have been reported females have out-performed males on the measure. Some researchers reporting gender differences favouring females have found these differences to be present exclusively in early adolescence and have therefore attributed these gender differences to differential maturation rates of males and females during puberty (e.g. Gibbs et al, 1992; Basinger, et al, 1995; Garmon et al 1996). A number of additional studies using the SRM-SF have also reported that adolescent females out-perform males during adolescence on the SRM-SF (e.g. Gregg et al, 1994; Palmer & Hollin, 2001; Krettenauer & Becker, 2001). However, a review of the literature reveals that "mapping" gender differences in SRM-SF scores onto known pubescent trends is problematic for a number of reasons.

Not all of the studies using the SRM-SF with both male and female participants have reported their findings regarding gender (examples of such studies include Ferguson et al, 1994; Palmer & Hollin, 1997; Krcmar & Valkenburg, 1999; Mizuno, 1999; Ferguson et al, 2001). The findings that do report gender differences conflict in terms of the ages at which gender differences favouring females emerge; furthermore, the broad age ranges of the participants in some studies - e.g. Gregg et al, 1994 (age range 13-18); Palmer & Hollin, 1998 (age range 13-22); Palmer & Hollin, 2001; (age range 12-18) - makes it difficult to interpret those gender differences that have been identified. The following examples should help to illustrate some of the difficulties associated with mapping the gender trends

of the SRM-SF: Humphries et al (2000) in their study of African Americans aged 10-13 found no gender differences favouring females; Ferguson & Cairns, 1996, found gender differences favouring females amongst 10-11 year olds, but not amongst 14-15 year olds; Krettenauer & Becker (2001) found gender differences favouring females amongst 16 year olds; Basinger et al (1995) found gender differences favouring females amongst 14 year olds but not amongst 10 year olds or 16-18 year olds; Comunian & Gielen found gender differences favouring females amongst 16 year olds and amongst 15-21 year olds (Comunian & Gielen 1995; Comunian & Gielen 2000); Barriga et al found no gender differences amongst 16-19 year olds (Barriga et al 2001). A study carried out by Mason & Gibbs in 1993 amongst college students in the U.S. identified gender differences that favoured older males; however, in a later study carried out by Van Ijzendoorn & Zwaart-Woudstra using a comparable sample, no effects for age or gender were found (Van Ijzendoorn & Zwaart-Woudstra, 1995).

The general finding that there is no apparent gender bias favouring males on terms of SRM-SF scores has been taken by many researchers as evidence to challenge Gilligan's (1982) hypothesis that males and females differ in terms of their use of Stage 3 and Stage 4 types of reasoning. However, it is by no means clear that it is appropriate to make this particular assumption on the basis of the published work. Typically, researchers using the SRM-SF have presented their findings in terms of summary mean scores over the questionnaire as a whole. However, use of summary mean scores alone may not be adequate when seeking to uncover patterns of Stage use amongst an adolescent sample. For example, if (as in some studies) the age range of adolescent sample is broad, and younger females are out-performing younger males, then any evidence of greater Stage 4 use in older males might be masked by comparatively greater use of the immature Stages by the younger boys. In any case, use of summary mean scores across a questionnaire may not be entirely representative indicators of adolescents' pattern of Stage use. It is clear that adolescence per se cannot not be represented as a fixed point that is half-way between "immaturity" and "maturity". One might, therefore, expect to see more variability in terms of Stage use within individuals and across questionnaire items amongst an adolescent sample than amongst children and adults. However, there are relatively few studies using the SRM-SF that have sought to investigate patterns of Stage use by adolescents across individual questionnaire items (two notable exceptions are Gregg et al, 1994 and Palmer & Hollin, 1998).

1.73 Differences between Offenders and Non-Offenders Identified by the SRM-SF

In line with findings using other measures of moral reasoning, researchers who have compared offending and non-offending adolescent samples using the SRM-SF typically report that offenders are "developmentally delayed" in terms of over-all moral reasoning with reference to their non-offending peers. Palmer & Hollin (1999) note that poor performance by offenders on moral reasoning tasks leaves the question of whether this is best explained by some form of cognitive structural deficit or whether it is because young offenders are "lagging behind" non-offenders in terms of moral development. They point out that this issue is critical when considering interventions for offenders. Clearly, in order to be able to provide a comprehensive answer to this question one needs to determine not only how this "lag" or "deficit" manifests itself among young people but also <u>when</u> in the developmental process it becomes evident. Furthermore, bearing in mind that in the UK at least, offending amongst males increases dramatically throughout the adolescent years, peaking at age 18, (see e.g. Coleman & Schofield 2001, p84) <u>early</u> intervention would seem to be key. However, curiously enough, the investigation of age related trends within adolescent offending samples is something that has received little attention in the SRM-SF research. In some studies, the age range of participant groups is broad; however, these groups have been treated as if they are homogenous groups of same-age adolescents.

1.74 Consensus Regarding "Importance" of Issues under Consideration

Few studies - two to the author's knowledge: Gregg, et al (1994); Palmer & Hollin, (1998) - have compared the relative "importance" ratings given by offender and non-offender samples to the initial evaluative component of each question (i.e. comparing the choices made from the given options of "very important" or "important" or "not important" in response to, say "How important is it to keep a promise to a friend?"). The two studies that have carried out this type of investigation have reported similar findings. Gregg, et al (1994) and Palmer & Hollin (1998) report that a high level of offenders and non-offenders - around 90% - evaluate most of the SRM-SF items as being "very important" or "important". In both studies, these findings are seen as providing general support for Kohlberg's proposal that certain moral values are universal (Gregg et al 1994, p551; Palmer & Hollin, 1998, p233) and that delinquents "recognised" (Palmer & Hollin 1998) or "acknowledged" (Gregg et al,

1994) the importance of these moral values. These interpretations seem rather odd (even if one disregards the controversy surrounding Kohlberg's claims regarding universal applicability of his moral typology). Firstly, it is difficult to see how generalisations about the "universality" of moral values can be made on the basis of these types of ratings given by two samples of Western adolescents. Secondly, it is not immediately apparent where the information that young offenders can "recognise" the importance of certain moral values leads in theoretical terms. Presumably, convicted young offenders have been made all too well aware of the relative "importance" that society attaches to e.g. law abiding behaviour through their contact with the criminal justice system. Finally, it is not clear from the studies that the researchers made reference to the content of the justifications when categorising responses as "important" or "not important". For example, it is possible that some participants circled the "important" response to an item but then went on to provide a disavowal in their justification (e.g. "it's important but....."). If this type of response was assigned to an "important" category, then this may have resulted in a distorted picture of the true level of consensus regarding the importance of the items. The evaluative component of the SRM-SF is, then, an aspect of the questionnaire that seems worthy of further investigation.

1.8 Overview to Empirical Research Presented in the Current Thesis

From the research review presented in this chapter it can be seen that there are a number of theoretical and methodological issues that need to be considered when seeking to investigate between-group differences in moral reasoning amongst adolescents. The development of the SRM-SF provides the researcher with a "user-friendly" tool that enables one to explore these issues in some depth.

The empirical research presented in this thesis naturally falls into two parts, and is presented in Chapter 2 and Chapter 3 of this thesis. In Chapter 2, work is presented that represents a partial replication of previous research using the SRM-SF to compare between-group differences amongst (male) offending and male and female non-offending adolescents (see e.g. Gregg et al, 1994; Palmer & Hollin, 1998). A key aim of this partial replication is to corroborate previous findings using "conventional" techniques to analyse SRM-SF data. In order to accomplish this, between-group comparisons are carried out with reference to summary SRM-SF mean scores. Claims regarding the

comparative moral immaturity of offenders in relation to their non-offending peers are investigated. Between-group differences amongst male and female non-offending adolescents are explored, however, a variant to previous work is provided in that the age-range of participants comes under particular scrutiny. In addition to carrying out between-group comparisons of the moral reasoning of participants the internal characteristics of the SRM-SF scale come under investigation.

The bulk of the "original contribution to research in the field" comes from the empirical studies that are presented in Chapter 3. As a result of the analyses presented in Chapter 2, issues are raised regarding the reliance on summary SRM-SF mean scores as the sole source of information to represent the moral reasoning of participants. In Chapter 3, SRM-SF data are analysed using alternatives to summary mean scores. Properties of the Global Moral Stage Status Scale (an alternative summary index presented by Gibbs et al (1992) to represent SRM-SF data) are investigated and issues regarding use of this scale are explored. The notion is put forward that through use of "categorical" techniques one may uncover between-group differences in patterns of SRM-SF stage use that were not revealed with reference to summary mean scores. It is noted that little attention has been paid in the SRM-SF research towards potential sources of intra-individual variation in patterns of moral stage use. As a result, potential sources of inter-individual and intra-individual variability in moral stage use are explored. The empirical research concludes with a demonstration of how recoding SRM-SF "mature stages" may impact on findings relating to gender differences in moral reasoning.

Chapter 4 summarises key findings that emerged from the empirical studies and discusses their theoretical and methodological implications. Suggestions are made for further research and concluding remarks are made.

CHAPTER 2: COMPARISON OF PATTERNS OF MORAL JUDGEMENT OF OFFENDING AND NON-OFFENDING ADOLESCENTS USING THE SRM-SF

2.1 Overview

In this chapter findings are presented resulting from work carried out using the SRM-SF to examine the differences in moral reasoning amongst male offending and male and female non-offending adolescents with reference to SRM-SF summary means. The first part of the chapter presents a description of the work to be carried out. In the second part of the chapter the results of the empirical findings are presented. In the final part of the chapter the findings are discussed in theoretical context and issues are raised that will be the subject of further investigation in Chapter 3.

2.2 Research aims for body of work presented in this chapter

The primary aim of the work carried out in this chapter is to replicate previous findings obtained using the SRM-SF to compare the moral reasoning of male offending and male and female non-offending adults. A further aim of the work to be carried out in this chapter is to test the internal characteristics of the SRM-SF scale. With that in mind, the specific aims of the empirical work and the key predictions that will be tested are as follows:

Tests will be carried out in order to attempt to replicated previous findings that have established that the SRM-SF is sensitive to age, and is therefore suitable for exploring developmental trends amongst young people. Comparisons will be carried out of SRM-SF summary means and items to test the claim that the SRM-SF effectively discriminates between offenders and non-offenders in that offending adolescents demonstrate comparative "moral immaturity" on the measure with reference to their non-offending peers. Investigation of gender differences on the measure will be carried out, in order to test the claim that the SRM-SF is not gender biased in terms of favouring males. Investigation of age related trends within the participant samples will be carried out. A specific aim is to test the prediction that the SRM-SF effectively discriminates between offenders and non-offenders across all age groups represented in the sample. Investigations of the internal characteristics of the SRM-SF will be carried out, in order to test the claim that the SRM-SF represents a "unitary scale" in terms of factor structure. Finally an investigation will be carried out about levels of consensus of the evaluation of the "importance" of the SRM-SF items amongst participants, the purpose of this

investigation is to test the claim that issues under consideration in the SRM-SF represent commonly held values amongst all participant groups.

2.3 Description of the SRM-SF (instrument used in the study):

The measure consists of 11 questions relating to the following "commonly held" sociomoral values: "contract and truth", "affiliation", "life", "property and law" and "legal justice". Each question is in two parts. Firstly, participants are asked to evaluate how important they think it is, say, to keep a promise to a friend - "very important", "important" or "not important". Secondly, participants are asked to justify why they think that is "very important", "important" or "not important". The 11 items in the questionnaire and the sociomoral values they relate to are as follows: (Questionnaire items are presented in full in Appendix 1

Contract and Truth:

Q1: Keeping a promise to a friend

Q2: Keep a promise to a stranger

Q3: Parents keeping promises to children

Q4: Telling the truth

Affiliation:

Q5: Children helping parents

Q6: Saving the life of a friend

Life:

Q7: Saving the life of a stranger

Q8: Living even if you don't want to

Property and Law:

Q9: Not taking things that belong to others

Q10: Obeying the law

Legal Justice:

Q11: Judges sending law-breakers to jail

2.3i Description of scoring procedure of the SRM-SF

Scoring the SRM-SF takes place by "matching" the moral justifications given to each questionnaire item to protocols in the SRM-SF handbook. Each justification is coded according to the major stage or transitional stage rating. For the purposes of statistical analysis, stage ratings are converted into scores ranging from 100 – 400. So, for example, a "Stage 1" response would be given a score of 100, a "Transition 2/3" response would be given a score of 250. The primary summary score used for analysis of the SRM-SF is the Sociomoral Reflection Maturity Score (SRMS), which is the mean of all scorable responses. It should be noted that certain types of responses are considered to be "unscorable"; these include tautologies, "word salads" and justifications where a single "response

unit" matches equally well with protocols from disparate moral stages. Questionnaires resulting in less than 7 out of 11 scorable items are deemed invalid and are discarded. Once the SRMS score has been calculated, a "Global Stage" status may also be assigned to the questionnaire. "Global Stage" is a ten-point scale that corresponds to "the developmental vicinity in which an SRMS is located" (Gibbs, Basinger and Fuller, 1992). Issues relating to "Global Moral Stage Status" ratings are taken up in Chapter 3 of this thesis.

2.4 METHOD

2.41 Participants

In total, 527 young people provided data for use in the current study. The participants were 190 male offenders, mean age 15.40 years (SD 1.24); 191 male non-offenders, mean age 14.97 years (SD 0.89); 146 female non-offenders, mean age 14.76 years (SD 0.90). The offending sample was drawn from 23 different settings in England (geographical range from Cornwall to County Durham). Of the 23 settings, 16 were Youth Offending Teams, 3 were Local Authority Secure Units, 2 were Secure Training Centres and 2 were Young Offender Institutions. Of the 190 offenders completing questionnaires, 95 were in custody and 95 were following community sentences under the supervision of a YOT. The non-offending sample was drawn from 6 secondary schools in England – 3 in the North of England, and 3 in the Greater London area. The majority of the non-offenders completing SRM questionnaires came from school year groups 9, 10 and 11.

2.42 Procedure

Participation was voluntary, and young people were assured that any information supplied would remain anonymous and confidential. Due to literacy issues, approximately 65% of the young offenders completed the tasks as interviews on a one to one basis; the questionnaire administrator writing down the responses given. The remainder of the young offenders and all the non-offenders completed the questionnaire as a pen-to-paper task.

2.43 Prior Experience of the Author in Coding SRM-SF Protocols and Inter-Rater Reliability

As noted earlier, the SRM-SF was designed in part as a more user-friendly alternative to the MJI, and initial self-training in use of the measure may be accomplished within approximately 30 hours. Self training in use of the measure is accomplished in three phases: overall familiarisation with summaries of the moral stages and aspects within stages; scoring practice by sample item responses; and scoring practice of sample questionnaires (Gibbs et al, 1992:45). In the manual, guidelines are given for self-study and practice exercises taken from 50 SRM-SF questionnaires with annotated answer keys are provided. The current author has completed the self-training as set out in the SRM-SF manual. Furthermore, she has obtained additional familiarisation with the SRM-SF through her involvement with a number of research projects conducted in UK schools that have used the SRM-SF as an assessment tool. This has resulted in the author having experience of coding upwards of 1000 SRM-SF questionnaires prior to analysing the data set used in the research project that is presented in the current thesis.

Aside from successfully completing of self-training, demonstrating competence in use of the measure requires that the trainee is able to achieve minimum standards of inter-rater reliability with an experienced SRM-SF rater. Whilst it was not possible to obtain indices of inter-rater reliability with the data set used in the current study, satisfactory inter-rater reliability had previously been demonstrated between the current author and another experienced SRM-SF protocol coder using a related data set. Gibbs, et al (1992) consider that in order to establish inter-rater reliability, raters should independently score 20-30 SRM-SF questionnaires. The minimum standards for acceptable inter-rater reliability are considered to be as follows: SRMS correlation: $r = .80$; mean absolute SRMS discrepancy = 0.20 points (this is equivalent to 20 points when stage ratings are converted by multiplying by 100 for the purposes of statistical analysis); global stage agreement within one GMSS interval: 80%; exact global stage agreement 50% (Gibbs et al, 1992: 56-57).

When evaluating inter-rater reliability on the related sample of 26 SRM-SF questionnaires, the current author and second experienced SRM-SF rater achieved the following standards: SRMS correlation:

n = 26, r=.847, p<0.001, two tailed test; mean absolute discrepancy = 0.138 points (or 13.8 points when multiplying stage ratings by 100); global stage agreement within one GMSS interval: 88%; exact global stage agreement 62%. It can be seen, then, that the standards achieved by the current author and the second rater more than adequately fulfilled the minimum requirements for inter-rater reliability on all criteria set out in the SRM-SF manual.

2.44 Treatment of Data

Typically, responses to the SRM-SF are scored questionnaire by questionnaire. For each participant, items 1 through 11 are coded, and summary moral reasoning scores calculated accordingly. There are certain advantages to treating the data in this way. Whilst extensive procedural guidelines are given in the SRM-SF manual for scoring questionnaires, scoring the SRM-SF is in part an inferential process (as is the case with any procedure involving textual analysis of open-ended questions). Moral justifications given by participants frequently do not match word-for-word with justifications given in the SRM-SF manual. Sometimes the wording of responses is ambiguous, and though the manual has guidelines for dealing with certain types of ambiguous responses (for example, where a moral justification appears to match equally well with protocols differing by two adjacent stage levels, the justification is scored at the higher level) these do not cover every eventuality. By referring other responses a participant has given, one may be able to get "more of a feel" for what the participant meant, and thus infer a stage rating. Treating the data in this way maximises the potential for internal consistency within participant questionnaires, and for building up a meaningful and valid moral profile for each individual. However, in terms of assessing global reliability of the SRM-SF, this method of treating the data is potentially problematic. For example, it raises the possibility that identical moral justifications by participants given to particular items are not given identical stage ratings. Furthermore, cues such as handwriting and spelling may unintentionally bias the rater towards reading in a greater or lower level of moral maturity from the questionnaire as a whole.
It was felt that for the purposes of the current study, a more rigorous approach that would be more testing to the overall reliability of the SRM-SF would be to score responses item by item and then calculate SRM-SF summary scores. Responses to SRM-SF questions were typed out into Microsoft Word tables, each question forming a separate Word table. Spelling mistakes were corrected and

some minimal editing regarding the syntax of items was carried out. For example, occurrences of "it's" and "it is" were standardised to "it's". No alterations were made to the semantics of any response. The tables for each question were sorted by text, and scored accordingly. This helped to ensure internal consistency in scoring, as identical or highly similar justifications were given identical stage ratings. All other aspects of scoring and summarising the data for Study 1 were carried out according to the procedure already described above.

2.45 "Unscorable" Items

As has been mentioned earlier, certain types of responses are classed as unscorable. Questionnaires resulting in more than 4/11 unscorable responses are excluded from analysis. Clearly, in a study of this nature, a certain amount of attrition is to be predicted simply due to non-completion of the questionnaire. All the non-offenders completed the task as pen-to-paper exercises in class and it was not possible to monitor for missing responses to items before the questionnaires were turned in. During interviews it was noted that some of the offenders were unable or unwilling to offer moral justifications to some of the items. Predictably, some participants gave responses that were ambiguous, or were not "moral justifications" as such – for example by repeating the question. Nevertheless, whilst scoring the items it was noted that 4 items in particular seemed to be generating a number of "unscorable" responses. The 4 items were: "Keeping a promise to a stranger", "Saving the life of a stranger", "Living even if you don't want to" and "Judges sending people to jail". The high level of unscorable responses to the items "Keeping a promise to a stranger" and "Saving the life of a stranger" can be partially explained by a number of participants responding "same as before". Both these items follow on from questions that ask about keeping a promise to a friend and saving the life of a friend. If one is scoring the questionnaire participant by participant, then according to the SRM-SF manual, one could assign the same stage rating to this type of response as given to the previous question. However, in the current study, the questionnaire was being treated item by item, so it was felt that this would be an inappropriate scoring strategy to adopt. Therefore, responses of this type were classified as unscorable. The problem with the item "how important is it for judges to send people who break the law to jail?" may have been partly to do with the wording of the question. At least 40 participants (or about 8% of the participant sample) answered this item with the response "it

depends" or "it depends what they've done". Without further justification, it is unclear how this type of response should be scored. The high proportion of unscorable responses to the question "How important is it to live even if you don't want to" was partially explained by a number of participants giving answers like: "it's up to them" or "it's their choice" or "it's their life". Approximately 60 participants – about 11% of the sample, responded to the question in this way. There is in fact a discrepancy in the SRM-SF manual relating to this type of response, which may reflect a theoretical issue. The response "it's their life" to this item is given as an example of a Stage 2, Aspect 3 (freedoms) response on page 100. However, on page 106, when giving examples of Transition Stage 3 / 4 reasoning, this response is cited as being "unscorable". It seems that this type of unelaborated response could correspond equally well to two disparate moral stages – Stage 2 (Aspect 3, "Freedoms") and Transition 3 / 4 (which incorporates the style of reasoning known as "Relativism of Personal Values).

2.46 Attrition Rate:

Attrition rate was relatively high – out of an initial 527 questionnaires that were turned in by participants, 413 met the initial criteria for inclusion in the preliminary analyses of the SRM-SF. The high attrition rate was in part due to the issues regarding scoring discussed above. However, it appeared that age was also a factor accounting for attrition rate, as the mean ages of participant groups completing scorable questionnaires were somewhat higher than for the participant groups as a whole. For male participants the difference between mean ages of the participant sample as a whole and those left after attrition was approximately 6 months. For the female sample the difference in mean ages was approximately 5 months.

The characteristics of participants providing "scorable" questionnaires according to the criteria were as follows: 147 male offenders, mean age 15.91 years, SD 1.30; 149 male non-offenders, mean age 15.45 years (SD 0.89); 117 female non-offenders, mean age 15.13 years (SD 0.89).

2.5 RESULTS

2.51. Comparison of SRM-SF Means by Participant Group:

The overall SRM-SF summary mean scores were as follows: offenders, mean = 222.59 (n=147, SD=30.11); male non-offenders mean = 264.36 (n=149, SD=44.79), female non-offenders mean = 262.48 (n=117, SD=29.88). Pearson correlations were carried out to test the prediction that there would be a positive relationship between age and mean scores on the SRM-SF. Results of 1-tailed Pearson correlations confirmed this finding for the relationship between the SRM-SF measure and age for the participant sample as a whole (n=413, r=.108, $p<0.05$) and for each of the participant groups (for the offenders, n=147 r=.170, $p<0.05$; for the male non-offenders, n=149, r=.418, $p<0.001$; for the female non-offenders, n=117, r=.278, $p<0.01$).

Statistical comparisons of means of offenders, male non-offenders and female non-offenders on the SRM-SF measure were carried out by ANCOVAs, using age as a covariate. Results of the ANCOVAs revealed significant differences between offenders and male non offenders $F(1,293) = 111.49$, $p<0.001$. The result of the ANCOVA comparing male and female non-offenders was not statistically significant: $F(1,263) = 0.50$, ns.

2.52. Classification of Participants into Age Categories

Participants were categorized according to age group categories as follows: Age Group 1: n = 103, mean age 14.13, SD 0.63; Age Group 2: n = 103, mean age 15.22, SD 0.20; Age Group 3: n = 207, mean age 16.37, SD 0.69. Initially 4 age group categories were created, corresponding to age quartiles – Age Group 1 and 2 as already described (being the lower two quartiles); Quartile 3, n=104, mean age 15.88, SD 0.17 and Quartile 4: n= 103 mean age 16.88, SD 0.62. However, due to age differences between offenders and non-offenders there were disproportionately high numbers of offenders in Quartile 4. Furthermore, due to differences in the sample size of male and female non-offenders, there were comparatively few female non-offenders that were assigned to this quartile. Therefore, for the purposes of analysis the upper quartiles were combined into one larger category to form Age Group 3. Broadly speaking, the 3 age group categories used for analysis represent the 3 school year groups from which the majority of the non-offending sample was drawn. Category 1 approximates to Year 9 (13 to 14 year olds), Category 2 to Year 10 (14 to 15 year olds), Category 3 to

Table 2.3: Summary Statistics for SRM-SF ANOVAs comparing between-group differences by Age Groups by participant type

		Age Group 1 & Age Group 2	Age Group 2 & Age Group 3	Age Group 1 & Age Group 3
	Participants	F (df)		
SRM-SF	Offenders	0.12 (1,64)	2.63 (1,112)	1.81 (1,112)
	Male Non-Offenders	10.11 (1,60)**	4.43 (1,115)*	35.51 (1,117)***
	Female Non-Offenders	2.92 (1,76)	1.86 (1,77)	11.23 (1,75)**

*p<0.05, **p<0.01, ***p<0.001

From Table 2.3 it can be seen that offender SRM-SF mean scores did not differ significantly across any of the age group categories. For male non-offenders significant differences in SRM-SF means were found between Age Groups 1 and 2 and also between Age Groups 2 and 3 (in both cases older participants having significantly higher means). For female non-offenders, significant differences in SRM-SF means were found between Age Groups 1 and 3 only (again, with older participants having significantly higher means).

2.54. Comparison of SRM-SF "socio-moral values" and individual items by participant group:

As described in the overview to the SRM-SF, the 11 questionnaire items represent the following sociomoral values: "Contract and Truth" (items 1 to 4), "Affiliation" (items 5 and 6), "Life" (items 7 and 8), "Property & Law" (items 9 and 10) and "Legal justice" (item 11). Using the participant sample as a whole, Pearson correlations were carried out to investigate the relationship between the sociomoral values and also to investigate the relationships between each of the individual questionnaire items. The prediction was that there would be positive relationships between the variables being compared. Results of 1-tailed Pearson correlations confirmed this for all the for all the inter-value correlations at $p<0.001$and for all the inter-item correlations at $p<0.05$ (the vast majority of the inter-item correlations achieved statistical significance at $p<0.001$). Correlation matrices giving Pearson r statistics and significance levels of the inter- variable comparisons may be found in Appendix 2. Having established that there were positive relationships between all the variables, ANCOVAs were carried out, using age as a covariate, to compare the means of participant groups on each of the SRM-SF sociomoral values and individual questionnaire items within each value. Summary tables showing descriptive statistics and the results of ANCOVAs are presented in Table 2.4 and 2.5 overleaf:

Table 2.4 SRM-SF sociomoral value and item means by participant group– descriptive statistics

	Offenders			♂ Non-Offenders			♀ Non-Offenders			Total		
Values & Items	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Contract & Truth	**147**	**228.40**	**39.84**	**149**	**277.57**	**49.94**	**116**	**271.88**	**37.43**	**412**	**258.42**	**48.58**
Promising Friend	*130*	*236.92*	*66.63*	*137*	*279.93*	*60.47*	*113*	*288.50*	*47.72*	*380*	*267.76*	*63.23*
Promising Stranger	*113*	*225.22*	*54.33*	*105*	*266.67*	*71.95*	*93*	*267.74*	*54.50*	*311*	*251.93*	*63.98*
Promising Children	*124*	*239.11*	*67.94*	*131*	*301.53*	*68.20*	*98*	*293.37*	*62.85*	*353*	*277.34*	*72.27*
Telling Truth	*133*	*218.80*	*45.89*	*124*	*266.13*	*66.43*	*99*	*237.37*	*56.85*	*356*	*240.45*	*60.11*
Affiliation	**144**	**233.16**	**52.04**	**147**	**264.63**	**59.02**	**115**	**272.17**	**37.58**	**406**	**255.60**	**53.81**
Help Parents	*129*	*234.50*	*55.83*	*130*	*268.85*	*60.77*	*103*	*269.90*	*43.35*	*362*	*256.91*	*56.87*
Saving Friend	*127*	*236.22*	*66.87*	*121*	*261.16*	*76.51*	*92*	*279.35*	*39.33*	*340*	*256.76*	*66.73*
Life	**134**	**239.37**	**41.63**	**129**	**260.08**	**60.56**	**102**	**269.12**	**48.64**	**365**	**255.00**	**52.29**
Saving Stranger	*95*	*238.95*	*57.05*	*96*	*265.10*	*64.12*	*65*	*256.15*	*56.26*	*256*	*253.13*	*60.47*
Staying Alive	*113*	*241.15*	*50.10*	*98*	*265.31*	*66.35*	*90*	*273.33*	*52.04*	*301*	*258.64*	*57.95*
Property & Law	**145**	**192.59**	**49.84**	**145**	**245.86**	**76.67**	**111**	**244.37**	**59.23**	**401**	**226.18**	**67.94**
Stealing	*126*	*194.05*	*57.83*	*125*	*216.80*	*81.80*	*98*	*231.63*	*63.97*	*349*	*212.75*	*70.42*
Obeying Law	*116*	*194.83*	*63.38*	*116*	*282.33*	*89.26*	*90*	*257.78*	*70.68*	*322*	*243.94*	*84.48*
Legal Justice	**104**	**195.67**	**58.75**	**115**	**249.57**	**86.47**	**86**	**228.49**	**59.18**	**305**	**225.25**	**74.02**

Table 2.5 SRM-SF sociomoral value and item means by participant group – ANCOVA statistics:

	Offender ♂ Non-Offender	♂ Non-Offender ♀ Non-Offender
Item	**F (df)**	**F (df)**
Contract & Truth	**101.27 (1,293) *****	**0.07 (1,262)**
Promise Friend	*35.30 (1,264) ****	*3.62 (1,247)*
Promise Stranger	*30.92 (1,215) ****	*0.62 (1,195)*
Promise Children	*59.27 (1,252) ****	*0.24 (1,226)*
Tell Truth	*48.11 (1,254) ****	*8.32 (1,220) ***
Affiliation	**35.17 (1,288) *****	**4.45 (1,259) ***
Help Parents	*28.44 (1,256) ****	*0.87 (1,230)*
Save Friend	*14.33 (1,245) ****	*7.60 (1,210) ***
Life	**15.91 (1,260) *****	**4.02 (1,228) ***
Save Stranger	*10.90 (1,188) ***	*0.20 (1,158)*
Stay Alive	*13.63 (1,208) ****	*2.22 (1,185)*
Property & Law	**54.43 (1,287) *****	**0.16 (1,253)**
Not Stealing	*8.25 (1,248) ***	*3.23 (1,220)*
Obey Law	*78.92 (1,229) ****	*2.17 (1,203)*
Legal Justice	**31.72 (1,216) *****	**1.76 (1,198)**

[a] Approaches statistical significance at $p<0.052$

From Table 2.5 it can be seen that male non-offenders demonstrated significantly higher mean scores than offenders on all socio-moral values and items. Differences in the mean scores of male and female non-offenders were significant for two of the values: "Affiliation" and "Life", with female non-offenders having significantly higher means than male non-offenders. Differences between male and female non-offenders were also found on two individual items: "How important is it to tell the truth?" and "How important is it to save the life of a friend?" On the item "How important is it to tell

the truth?" male non-offenders demonstrated significantly higher means than female non-offenders. On the item "How important is it to save the life of a friend?" female non-offenders demonstrated significantly higher means than male non-offenders.

2.55. SRM-SF socio-moral value and item means by participant group and age:

SRM-SF item means for participant groups in each of the 3 age categories are summarised in Table 2.6 overleaf. In line with analyses carried out for the overall SRM-SF means, two types of ANOVAs were carried out. Firstly, ANOVAs were carried out to compare the means between participant groups on SRM-SF socio-moral values and items within each of the 3 age group categories. Secondly, ANOVAs were carried out to compare the means across age group categories within participant groups. Summary statistics showing the results of the ANOVAs are presented in Tables 2.7 and 2.8 on pages 45 & 46. Key findings of note from the analyses are described on page 47 following the tables.

Table 2.6: Descriptive statistics for SRM-SF socio-moral values and items across age groups

	Offenders			♂ Non-Offenders			♀ Non-Offenders			All Participants		
Values and Items	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
Contract & Truth Age Group 1	**33**	**225.38**	**32.45**	**32**	**241.02**	**51.27**	**37**	**260.36**	**42.75**	**102**	**242.97**	**44.78**
Age Group 2	**33**	**227.40**	**38.61**	**30**	**277.78**	**43.40**	**40**	**275.94**	**32.24**	**103**	**260.92**	**44.00**
Age Group 3	**81**	**230.04**	**43.26**	**87**	**290.95**	**45.07**	**39**	**278.63**	**35.33**	**207**	**264.79**	**51.04**
Promise Friend Age Group 1	*29*	*234.48*	*73.32*	*29*	*236.21*	*66.68*	*36*	*281.94*	*53.66*	*94*	*253.19*	*67.52*
Age Group 2	*30*	*246.67*	*45.36*	*29*	*279.31*	*59.04*	*39*	*291.03*	*44.24*	*98*	*273.98*	*52.38*
Age Group 3	*71*	*233.80*	*71.60*	*79*	*296.20*	*50.49*	*38*	*292.11*	*45.81*	*188*	*271.81*	*65.43*
Promise Stranger Age Group 1	*25*	*218.00*	*47.61*	*22*	*220.45*	*59.08*	*29*	*256.90*	*59.35*	*76*	*233.55*	*57.96*
Age Group 2	*20*	*207.50*	*56.84*	*24*	*258.33*	*52.48*	*36*	*261.11*	*46.46*	*80*	*246.88*	*55.31*
Age Group 3	*68*	*233.09*	*55.05*	*59*	*287.29*	*75.20*	*28*	*287.50*	*55.49*	*155*	*263.55*	*68.69*
Promise Children Age Group 1	*28*	*233.93*	*57.82*	*25*	*266.00*	*68.80*	*31*	*279.03*	*68.04*	*84*	*260.12*	*67.12*
Age Group 2	*30*	*236.67*	*78.71*	*26*	*303.85*	*67.71*	*34*	*301.47*	*52.92*	*90*	*280.56*	*72.94*
Age Group 3	*66*	*242.42*	*67.51*	*80*	*311.88*	*65.24*	*33*	*298.49*	*66.71*	*179*	*283.80*	*73.36*
Tell Truth Age Group 1	*32*	*218.75*	*27.68*	*23*	*236.96*	*48.19*	*32*	*218.75*	*61.89*	*87*	*223.56*	*48.13*
Age Group 2	*32*	*212.50*	*47.52*	*24*	*264.58*	*63.38*	*35*	*251.43*	*56.21*	*91*	*241.21*	*58.97*
Age Group 3	*69*	*221.74*	*51.79*	*77*	*275.32*	*70.07*	*32*	*240.63*	*48.26*	*178*	*248.31*	*64.42*
Affiliation Age Group 1	**33**	**224.24**	**55.02**	**32**	**225.78**	**57.67**	**37**	**266.22**	**26.50**	**102**	**239.95**	**51.25**
Age Group 2	**32**	**221.88**	**54.53**	**30**	**250.83**	**53.53**	**39**	**271.79**	**46.66**	**101**	**249.75**	**54.94**
Age Group 3	**79**	**241.46**	**48.85**	**85**	**284.12**	**53.01**	**39**	**278.21**	**36.36**	**203**	**266.38**	**52.35**
Help Parents Age Group 1	*32*	*223.44*	*55.34*	*25*	*236.00*	*36.86*	*34*	*260.29*	*36.47*	*91*	*240.66*	*46.50*
Age Group 2	*28*	*225.00*	*39.67*	*27*	*274.07*	*64.12*	*34*	*270.59*	*46.25*	*89*	*257.30*	*54.64*
Age Group 3	*69*	*243.48*	*60.58*	*78*	*277.56*	*62.75*	*35*	*278.57*	*45.83*	*182*	*264.84*	*61.07*
Save Friend Age Group 1	*26*	*228.85*	*75.06*	*26*	*215.38*	*83.39*	*31*	*275.81*	*36.22*	*83*	*242.17*	*70.92*
Age Group 2	*28*	*223.21*	*79.91*	*27*	*233.33*	*69.34*	*31*	*275.81*	*48.08*	*86*	*245.35*	*69.72*
Age Group 3	*73*	*243.84*	*57.70*	*68*	*289.71*	*63.81*	*30*	*286.67*	*31.98*	*171*	*269.59*	*60.72*
Life Age Group 1	**30**	**231.67**	**34.07**	**30**	**230.00**	**64.46**	**33**	**250.76**	**44.85**	**93**	**237.90**	**49.67**
Age Group 2	**28**	**234.82**	**38.09**	**23**	**251.09**	**63.73**	**35**	**277.14**	**47.89**	**86**	**256.40**	**52.61**
Age Group 3	**76**	**244.08**	**45.25**	**76**	**274.67**	**53.46**	**34**	**278.68**	**49.29**	**186**	**262.90**	**51.68**
Save Stranger Age Group 1	*20*	*242.50*	*40.64*	*21*	*245.24*	*72.29*	*22*	*240.91*	*45.35*	*63*	*242.86*	*53.77*
Age Group 2	*18*	*225.00*	*52.16*	*19*	*242.12*	*65.11*	*23*	*276.09*	*42.29*	*60*	*250.00*	*56.75*
Age Group 3	*57*	*242.11*	*63.24*	*56*	*280.36*	*56.95*	*20*	*250.00*	*74.34*	*133*	*259.40*	*64.59*
Stay Alive Age Group 1	*28*	*226.79*	*53.55*	*26*	*221.15*	*61.92*	*29*	*260.34*	*52.41*	*83*	*236.75*	*57.96*
Age Group 2	*22*	*247.73*	*47.50*	*15*	*286.67*	*61.14*	*30*	*273.33*	*55.29*	*67*	*267.91*	*55.53*
Age Group 3	*63*	*245.24*	*48.95*	*57*	*279.82*	*61.13*	*31*	*285.49*	*46.89*	*151*	*266.56*	*56.19*
Property & Law Age Group 1	**32**	**186.72**	**43.53**	**32**	**226.56**	**68.38**	**38**	**228.29**	**41.98**	**102**	**214.71**	**54.99**
Age Group 2	**33**	**181.06**	**51.17**	**30**	**239.17**	**66.53**	**36**	**237.50**	**67.48**	**99**	**219.19**	**67.27**
Age Group 3	**80**	**199.69**	**51.02**	**83**	**255.72**	**82.02**	**37**	**267.57**	**60.05**	**200**	**235.50**	**73.04**
Not Stealing Age Group 1	*28*	*192.86*	*53.94*	*31*	*198.39*	*79.04*	*35*	*228.57*	*53.26*	*94*	*207.98*	*64.47*
Age Group 2	*26*	*188.46*	*58.83*	*25*	*216.00*	*64.10*	*29*	*222.41*	*71.45*	*80*	*209.38*	*66.13*
Age Group 3	*72*	*196.53*	*59.53*	*69*	*225.36*	*88.13*	*34*	*242.65*	*67.57*	*175*	*216.86*	*75.34*
Obey Law Age Group 1	*26*	*186.54*	*48.08*	*27*	*250.00*	*88.80*	*31*	*225.81*	*58.98*	*84*	*221.43*	*71.26*
Age Group 2	*27*	*181.48*	*66.72*	*24*	*281.25*	*85.74*	*31*	*253.23*	*75.21*	*82*	*237.81*	*85.55*
Age Group 3	*63*	*203.97*	*66.78*	*65*	*296.15*	*88.52*	*28*	*298.21*	*58.50*	*156*	*259.29*	*87.77*
Legal Justice Age Group 1	**27**	**211.11**	**54.30**	**25**	**198.00**	**78.37**	**28**	**230.36**	**53.30**	**80**	**213.75**	**63.13**
Age Group 2	**25**	**180.00**	**35.36**	**26**	**248.08**	**78.08**	**31**	**212.90**	**60.51**	**82**	**214.02**	**65.86**
Age Group 3	**52**	**195.19**	**68.07**	**64**	**270.31**	**85.32**	**27**	**244.44**	**60.98**	**143**	**238.11**	**82.00**

Table 2.7: Summary ANOVA Statistics for SRM-SF values and items between participants within Age Groups:

	Offender ♂ Non-Offender		♂ Non-Offender ♀ Non-Offender	
Item	**F (df)**		**F (df)**	
Contract & Truth Age Group 1	**2.17 (1,63)**		**2.92 (1,67)**	
Age Group 2	**23.78 (1,61)**	***	**0.04 (1,68)**	
Age Group 3	**79.62 (1,166)**	***	**2.28 (1,124)**	
Promise Friend Age Group 1	*0.01 (1,56)*		*9.40 (1,63)*	**
Age Group 2	*5.70 (1,57)*	*	*0.88 (1,66)*	
Age Group 3	*38.64 (1,148)*	***	*0.18 (1,115)*	
Promise Stranger Age Group 1	*0.03 (1,45)*		*4.74 (1,49)*	*
Age Group 2	*9.49 (1,42)*	**	*0.05 (1,58)*	
Age Group 3	*21.85 (1,125)*	***	*0.00 (1,85)*	
Promise Children Age Group 1	*3.40 (1,51)*		*0.50 (1,54)*	
Age Group 2	*11.54 (1,54)*	**	*0.02 (1,58)*	
Age Group 3	*39.72 (1,144)*	***	*0.97 (1,111)*	
Tell Truth Age Group 1	*3.14 (1,53)*		*1.38 (1,53)*	
Age Group 2	*12.37 (1,54)*	**	*0.70 (1,57)*	
Age Group 3	*27.09 (1,144)*	***	*6.54 (1,107)*	*
Affiliation Age Group 1	**0.01 (1,63)**		**14.64 (1,67)**	***
Age Group 2	**4.44 (1,60)**	*	**3.01 (1,67)**	
Age Group 3	**28.60 (1,162)**	***	**0.40 (1,122)**	
Help Parents Age Group 1	*0.96 (1,55)*		*6.34 (1,57)*	*
Age Group 2	*11.75 (1,53)*	**	*0.06 (1,59)*	
Age Group 3	*11.16 (1,145)*	**	*0.01 (1,111)*	
Save Friend Age Group 1	*0.37 (1,50)*		*13.32 (1,55)*	**
Age Group 2	*0.25 (1,53)*		*7.50 (1,56)*	**
Age Group 3	*20.09 (1,139)*	***	*0.06 (1,96)*	
Life Age Group 1	**0.02 (1,58)**		**2.23 (1,61)**	
Age Group 2	**1.27 (1,60)**		**3.15 (1,56)**	
Age Group 3	**14.50 (1,150)**	***	**0.14 (1,108)**	
Save Stranger Age Group 1	*0.02 (1,39)*		*0.06 (1,41)*	
Age Group 2	*0.77 (1,35)*		*4.16 (1,40)*	*
Age Group 3	*11.40(1,111)*	**	*3.55 (1,74)*	
Stay Alive Age Group 1	*0.13 (1,52)*		*6.46 (1,53)*	*
Age Group 2	*4.75 (1,35)*	*	*0.54 (1,43)*	
Age Group 3	*11.80 (1,118)*	**	*0.20 (1,86)*	
Property & Law Age Group 1	**7.73 (1,62)**	**	**0.02 (1,68)**	
Age Group 2	**15.26 (1,61)**	***	**0.01 (1,62)**	
Age Group 3	**27.19 (1,161)**	***	**0.62 (1,118)**	
Not Stealing Age Group 1	*0.10 (1,57)*		*3.38 (1,64)*	
Age Group 2	*2.56 (1,49)*		*0.12 (1,52)*	
Age Group 3	*5.22 (1,126)*	*	*1.01 (1,101)*	
Obey Law Age Group 1	*10.35 (1,51)*	**	*1.53 (1,56)*	
Age Group 2	*21.76 (1,49)*	***	*1.66 (1,53)*	
Age Group 3	*44.03 (1,126)*	***	*0.01 (1,91)*	
Legal Justice Age Group 1	**0.50 (1,50)**		**3.15 (1,51)**	
Age Group 2	**15.87 (1,49)**	***	**3.67 (1,55)**	
Age Group 3	**26.56 (1,114)**	***	**2.04 (1,89)**	

Table 2.8: Summary ANOVA statistics for SRM-SF values and items between Age Groups within participant samples:

		Age Group Comparisons		
		1 & 2	2 & 3	1 & 3
		F(df)	F (df)	F (df)
Contract & Truth	Offenders	**0.05 (1,57)**	**0.09 (1,112)**	**0.31 (1,112)**
	♂ Non-Offenders	**9.22 (1,60)** **	**1.94 (1,115)**	**26.64 (1,117)** ***
	♀ Non-Offenders	**3.29 (1,75)**	**0.13 (1,77)**	**4.14 (1,74)** *
Promise Friend	Offenders	*0.59 (1,57)*	*0.83 (1,99)*	*0.00 (1,98)*
	♂ Non-Offenders	*6.79 (1,56)* *	*2.16 (1,106)*	*25.03 (1,106)* ***
	♀ Non-Offenders	*0.64 (1,73)*	*0.01 (1,75)*	*0.77 (1,72)*
Promise Stranger	Offenders	*0.46 (1,43)*	*3.29 (1,86)*	*1.47 (1,91)*
	♂ Non-Offenders	*5.30 (1,44)* *	*2.96 (1,81)*	*14.09 (1,79)* ***
	♀ Non-Offenders	*0.10 (1,63)*	*4.29 (1,62)* *	*4.04 (1,55)* *
Promise Children	Offenders	*0.02 (1,56)*	*0.14 (1,94)*	*0.34 (1,92)*
	♂ Non-Offenders	*3.92 (1,49)* [a]	*0.29 (1,104)*	*9.18 (1,103)* **
	♀ Non-Offenders	*2.22 (1,63)*	*0.04 (1,65)*	*1.33(1,62)*
Tell Truth	Offenders	*0.41 (1,62)*	*0.73 (1,99)*	*0.09 (1,99)*
	♂ Non-Offenders	*2.81 (1,45)*	*0.45 (1,99)*	*6.02 (1,98)* *
	♀ Non-Offenders	*5.13 (1,65)* *	*0.71 (1,65)*	*2.49 (1,62)*
Affiliation	Offenders	**0.03 (1,63)**	**3.42 (1,109)**	**2.68 (1,110)**
	♂ Non-Offenders	**3.13 (1,60)**	**8.70 (1,113)** **	**28.83 (1,115)** ***
	♀ Non-Offenders	**0.41 (1,74)**	**0.46 (1,76)**	**2.68 (1,74)**
Help Parents	Offenders	*0.02 (1,58)*	*2.21 (1,95)*	*2.52 (1,99)*
	♂ Non-Offenders	*6.75 (1,50)* *	*0.06 (1,103)*	*9.84 (1,101)* **
	♀ Non-Offenders	*1.04 (1,66)*	*0.52 (1,67)*	*3.35 (1,67)*
Save Friend	Offenders	*0.07 (1,52)*	*2.07 (1,99)*	*1.10 (1,97)*
	♂ Non-Offenders	*0.73 (1,51)*	*14.36 (1,93)* ***	*21.40 (1,92)* ***
	♀ Non-Offenders	*0.00 (1,60)*	*1.07 (1,59)*	*1.54 (1,59)*
Life	Offenders	**0.11 (1,56)**	**0.93 (1,102)**	**1.84 (1,104)**
	♂ Non-Offenders	**1.41 (1,51)**	**3.14 (1,97)**	**13.33 (1,104)** ***
	♀ Non-Offenders	**5.48 (1,66)** *	**0.02 (1,67)**	**5.87 (1,65)** *
Save Stranger	Offenders	*1.35 (1,36)*	*1.08 (1,73)*	*0.00 (1,75)*
	♂ Non-Offenders	*0.02 (1,38)*	*5.95 (1,73)* *	*4.99 (1,75)* *
	♀ Non-Offenders	*7.25 (1,43)* *	*2.07 (1,41)*	*0.23 (1,40)*
Stay Alive	Offenders	*2.08 (1,48)*	*0.04 (1,83)*	*2.60 (1,89)*
	♂ Non-Offenders	*10.74 (1,39)* **	*0.15 (1,70)*	*16.31 (1,81)* ***
	♀ Non-Offenders	*0.86 (1,57)*	*0.86 (1,59)*	*3.84 (1,58)* [c]
Property & Law	Offenders	**0.23 (1,63)**	**3.11 (1,111)**	**1.60 (1,110)**
	♂ Non-Offenders	**0.54 (1,60)**	**0.99 (1,111)**	**3.19 (1,113)**
	♀ Non-Offenders	**0.50 (1,72)**	**4.05 (1,71)** *	**10.83 (1,73)** **
Not Stealing	Offenders	*0.08 (1,52)*	*0.35 (1,96)*	*0.08 (1,98)*
	♂ Non-Offenders	*0.81 (1,54)*	*0.24 (1,92)*	*2.13 (1,98)*
	♀ Non-Offenders	*0.16 (1,62)*	*1.33 (1,61)*	*0.93 (1,67)*
Obey Law	Offenders	*0.10 (1,51)*	*2.14(1,88)*	*1.46 (1,87)*
	♂ Non-Offenders	*1.63 (1,49)*	*0.51 (1,87)*	*5.18 (1,90)* *
	♀ Non-Offenders	*2.55 (1,60)*	*6.48 (1,57)* *	*22.34 (1,57)* ***
Legal Justice	Offenders	**5.89 (1,50)** * (-ve)	**1.10 (1,75)**	**1.11 (1,77)**
	♂ Non-Offenders	**5.22 (1,49)** *	**1.32 (1,88)**	**13.50 (1,87)** ***
	♀ Non-Offenders	**1.37 (1,57)**	**3.89 (1,56)** [b]	**0.83 (1,53)**

[a, b] Both approach statistical significance at p = 0.053

[c] Approaches statistical significance at p =0.055

2.55i. Differences in SRM-SF Value and Item Means between Participants within Age Groups

In Age Group 1, none of the differences between offenders and male non-offenders were statistically significant except for the "Property and Law" value and the individual item "Obeying the law" – where male non-offenders had significantly higher means than offenders. Female non-offenders had significantly higher means than male non-offenders on the "Affiliation" value and on five of the eleven individual questionnaire items ("Keeping a promise to a friend", "Keeping a promise to a stranger", "Helping parents", "Saving the life of a friend" and living even if you don't want to). In Age Group 2, male non-offenders demonstrated significantly higher means than offenders on four of the five sociomoral values (All those except "Life") and on eight of the eleven individual questionnaire items (all those except "Saving the life of a friend", "Saving the life of a stranger" and "Not taking things that belong to others"). In this age group, differences between male and female non-offenders were statistically significant for one individual item only: "Saving the life of a friend" where female non-offenders demonstrated significantly higher means than male non-offenders. In Age Group 3, male non-offenders had significantly higher mean scores than offenders on all sociomoral values and all individual items save for "How important is it not to Steal?" No significant differences between male and female non-offenders were found except for the item "Telling the truth". On this item male non-offenders demonstrated significantly higher mean scores than female non-offenders.

2.55ii. Differences in SRM-SF Value and Item Means across Age Groups by Participant Type

From Table 2.8 it can be seen that for offenders, between age-group differences were not significant for any of the sociomoral values or individual items except for differences in the means of the "Legal Justice" value between Age Groups 1 and 2. For this value, means in Age group 2 were significantly lower than those in Age Group 1. For male non-offenders, significant between age-group differences were found between Age Groups 1 and 2 for the "Life" and "Legal Justice" values (in both cases older participants having higher means) and 5/11 individual questionnaire items ("Keeping a promise to a friend", "Keeping a promise to a stranger", "Helping parents", "Staying alive" and "judges sending people to jail". Male non-offender mean scores differed significantly between Age Groups 2 and 3 for the "Affiliation" value and for the items "Saving the life of a stranger" and "living even if

you don't want to" (with means in age Group 3 being significantly higher than those in Age Group 2). Male non-offender means differed significantly between Age Groups 1 and 3 for all of the values except "Property and Law" and for all items except "not taking things that belong to others". For female non-offenders significant differences were found between Age Groups 1 and 2 for one value only - "Life" (females in Age Group 2 having higher means than those in Age Group 1) and for the items "Telling the truth" and "Saving the life of a stranger" (in both cases females in Age Group 2 had higher means than those in Age Group 1). Differences between Age Groups 2 and 3 were found for the value "Property and Law" and for the items "Keeping a promise to a stranger" and "obeying the law" (with females in Age Group 3 having higher means than those in Age Group 2). Female non-offender means differed between Age Group 1 and Age Group 3 on the values "Contract and Truth", "Life" and "Property and Law" and also on the items "Keeping a promise to strangers" and "Obeying the law" (females in Age Group 3 having higher means on these variables than those in Age Group 1).

2.56. Comparisons between Values and Intra-Value Items within the SRM-SF, by Participant Group

Paired t-tests were carried out to compare mean differences in scores between the SRM-SF socio-moral values and also between the intra-value item comparisons for each of the participant groups. Summary statistics for the related t-tests groups are presented in Tables 2.9 and 2.10. Key findings of note from the analyses are described on page 50, following the tables.

Table 2.9 Summary statistics from SRM-SF inter-value comparisons using paired t-tests:

Value Comparisons	Participants	t	df	Sig.
Contract and Truth / Affiliation	Offenders	-1.078	143	ns
	Male Non-Offenders	2.919	146	p<0.01
	Female Non-Offenders	-.185	113	ns
Contract and Truth / Life	Offenders	-2.114	133	p<0.05
	Male Non-Offenders	3.446	128	p<0.01
	Female Non-Offenders	.721	100	ns
Contract and Truth / Property and Law	Offenders	8.004	144	p<0.001
	Male Non-Offenders	5.195	144	p<0.001
	Female Non-Offenders	4.929	109	p<0.001
Contract and Truth / Legal Justice	Offenders	5.852	103	p<0.001
	Male Non-Offenders	3.444	114	p<0.01
	Female Non-Offenders	6.260	84	p<0.001
Affiliation / Life	Offenders	-1.333	130	ns
	Male Non-Offenders	1.203	127	ns
	Female Non-Offenders	1.229	99	ns
Affiliation / Property and Law	Offenders	7.823	141	p<0.001
	Male Non-Offenders	2.625	142	p<0.01
	Female Non-Offenders	4.430	108	p<0.001
Affiliation / Legal Justice	Offenders	5.753	101	p<0.001
	Male Non-Offenders	1.853	112	ns
	Female Non-Offenders	6.234	83	p<0.001
Life / Property and Law	Offenders	8.979	131	p<0.001
	Male Non-Offenders	.972	124	ns
	Female Non-Offenders	3.109	95	p<0.01
Life / Legal Justice	Offenders	6.742	94	p<0.001
	Male Non-Offenders	1.091	96	ns
	Female Non-Offenders	5.026	71	p<0.001
Property and Law / Legal Justice	Offenders	-1.035	103	ns
	Male Non-Offenders	.000	111	ns
	Female Non-Offenders	2.006	81	p<0.05

Table 2.10: Summary statistics from SRM-SF intra-value item comparisons using paired t-tests:

Values	Item Comparisons	Participants	t	df	Sig.
Contract and Truth	Promise Friend / Promise Stranger	Offenders	1.555	102	Ns
		Male Non-Offenders	2.370	99	p<0.05
		Female Non-Offenders	3.655	90	p<0.001
	Promise Friend / Promise Children	Offenders	-0.174	109	Ns
		Male Non-Offenders	-3.502	119	p<0.01
		Female Non-Offenders	-1.062	94	Ns
	Promise Friend / Tell Truth	Offenders	2.639	117	p<0.01
		Male Non-Offenders	2.176	115	p<0.05
		Female Non-Offenders	6.627	95	p<0.001
	Promise Stranger / Promise Children	Offenders	-1.422	97	Ns
		Male Non-Offenders	-4.244	91	p<0.001
		Female Non-Offenders	-2.973	76	p<0.01
	Promise Stranger / Tell Truth	Offenders	0.686	101	Ns
		Male Non-Offenders	-0.213	87	Ns
		Female Non-Offenders	3.330	77	p<0.01
	Promise Children / Tell Truth	Offenders	2.841	109	p<0.01
		Male Non-Offenders	4.959	110	p<0.001
		Female Non-Offenders	6.088	83	p<0.001
Affiliation	Help Parents / Save Friend	Offenders	-0.186	111	Ns
		Male Non-Offenders	0.614	103	Ns
		Female Non-Offenders	-2.476	79	p<0.05
Life	Save Stranger / Stay Alive	Offenders	-0.351	73	Ns
		Male Non-Offenders	-1.284	64	Ns
		Female Non-Offenders	-2.261	52	p<0.05
Property and Law	Not Stealing / Obeying Law	Offenders	-0.063	96	Ns
		Male Non-Offenders	-5.641	95	p<0.001
		Female Non-Offenders	-2.691	76	p<0.01

2.56i. Findings from SRM-SF inter-value comparisons:

For the offenders, paired comparisons of the SRM-SF socio-moral values revealed the following findings: "Contract and Truth had statistically higher means than "Life", "Property and Law" and "Legal Justice"; "Affiliation" had statistically higher means than "Property and Law" and "Legal Justice"; "Life" had statistically higher means than "Property and Law" and "Legal Justice". None of the other inter-value comparisons were statistically significant. Within the value of "Contract and Truth" the items "Keeping a promise to a friend" and "Keeping a promise to a child" both had significantly higher means than the item "Telling the truth". For the male non-offenders, paired comparisons of the SRM-SF socio-moral values revealed the following findings: "Contract and Truth" had significantly higher means than all other values under consideration; "Affiliation" had significantly higher means than "Property and Law". None of the other paired inter-value comparisons were statistically significant. Within the value of "Contract and Truth" the item "keeping a promise to a friend" had significantly higher means than all other intra-value items and the item "Keeping a promise to children had significantly higher means than "keeping a promise to a stranger". Within the "Property and Law" value, the item "Obeying the law had significantly higher means than the item "not taking things that belong to others". For the female non-offenders paired comparisons of the SRM-SF socio-moral values revealed the following findings: "Contract and Truth", "Affiliation" and "Life" all had significantly higher means than the "Property and Law" and "Legal Justice" values, the "Property and Law value had significantly higher means than the "Legal Justice" value. Within the "Contract and Truth" value "Telling the truth" had significantly lower means than all other items, "Keeping a promise to a friend" and "Keeping a promise to a child" had significantly higher means than "Keeping a promise to a stranger". Within the "Affiliation" value, "Saving the life of a friend" had significantly higher means than "Helping parents", within the "Life" value "Living even if you don't want to" had significantly higher means than "saving the life of a stranger" and within the "Property and Law" value "Obeying the law" had significantly higher means than "Not taking things that belong to others".

2.57 Factor Analyses of SRM-SF items

Some previous researchers have carried out investigations of the internal structure of the SRM-SF through the use of reliability analyses and Factor Analysis. Due to the high attrition rate on some of the SRM-SF items in this sample it was inappropriate to carry out a reliability analysis, as this would entail replacing a substantial number of missing values in the scale (please see the discussion for elaboration of this point).

Investigation of the internal structure of the SRM-SF was carried out for the whole sample by Factor analysis using Alpha factoring (excluding missing values pair-wise) with Varimax rotation. This technique extracted two factors with Eigenvalues greater than 1.000 (Eigenvalue for Factor 1= 4.129 accounting for 37.54% of variance; Eigenvalue for Factor 2=1.051, accounting for 9.56% of the variance). This technique was repeated for the non-offending sample only, also resulting in the extraction of two factors with Eigenvalues over 1.000 (Eigenvalue for Factor 1=3.900, accounting for 35.46% variance; Eigenvalue for Factor 2=1.107, accounting for 10.06% variance). Finally, the technique was carried out for the male sample only (offenders and non-offenders) Once again the technique extracted two factors with Eigenvalues over 1.000 (Eigenvalue 1 = 4.362, accounting for 39.65% of the variance, Eigenvalue 2 = 1.043, accounting for 9.48% of variance). Item loadings on the factors for each of the three analyses are presented in Tables 2.11 –2.13 below. Interpretation of the factors follows the tables.

Table 2.11 Item loadings for factor analysis carried out amongst the participant sample as a whole:

Factor 1		Factor 2	
Item	Load	Item	Load
Obeying the law	**.69**	**Keeping a promise to a stranger**	**.72**
Telling the truth	**.60**	**Keeping a promise to a friend**	**.59**
Judges sending people to jail	**.56**	**Not taking things that belong to others**	**.46**
Living even if you don't want to	**.48**	**Saving the life of a friend**	**.45**
Helping parents	**.43**	**Keeping a promise to a child**	**.41**
Keeping a promise to a child	**.42**	*Helping parents*	*.37*
Saving the life of a friend	*.35*	*Judges sending people to jail*	*.32*
Saving the life of a stranger	*.33*	*Saving the life of a stranger*	*.26*
Keeping a promise to a friend	*.24*	*Living even if you don't want to*	*.25*
Keeping a promise to a stranger	*.23*	*Obeying the law*	*.23*
Not taking things that belong to others	*.19*	*Telling the truth*	*.17*

Table 2.12: Item loadings for factor analysis carried out amongst non-offenders only:

Factor 1		Factor 2	
Item	Load	Item	Load
Keeping a promise to a stranger	**.68**	**Obeying the law**	**.85**
Saving the life of a friend	**.57**	**Judges sending people to jail**	**.54**
Keeping a promise to a friend	**.53**	**Telling the truth**	**.44**
Keeping a promise to a child	**.46**	*Saving the life of a stranger*	*.32*
Helping parents	**.44**	*Living even if you don't want to*	*.32*
Saving the life of a stranger	**.44**	*Helping parents*	*.31*
Not taking things that belong to others	**.42**	*Keeping a promise to a child*	*.29*
Living even if you don't want to	**.41**	*Keeping a promise to a friend*	*.23*
Judges sending people to jail	*.30*	*Not taking things that belong to others*	*.19*
Telling the truth	*.29*	*Saving the life of a friend*	*.17*
Obeying the law	*.12*	*Keeping a promise to a stranger*	*.11*

Table 2.13: Item loadings for factor analysis carried out amongst males only:

Factor 1		Factor 2	
Item	Load	Item	Load
Obeying the law	**.66**	**Keeping a promise to a stranger**	**.71**
Judges sending people to jail	**.64**	**Keeping a promise to a friend**	**.62**
Telling the truth	**.59**	**Not taking things that belong to others**	**.50**
Living even if you don't want to	**.54**	*Keeping a promise to a child*	*.39*
Helping parents	**.48**	*Saving the life of a friend*	*.39*
Keeping a promise to a child	**.46**	*Judges sending people to jail*	*.33*
Saving the life of a friend	*.39*	*Helping parents*	*.31*
Saving the life of a stranger	*.37*	*Obeying the law*	*.24*
Keeping a promise to a stranger	*.31*	*Telling the truth*	*.24*
Keeping a promise to a friend	*.25*	*Saving the life of a stranger*	*.21*
Not taking things that belong to others	*.19*	*Living even if you don't want to*	*.18*

Broadly speaking, interpretation of the factors seems to indicate some discrimination between a "truth, law and justice" domain and a "life and relationships" domain. The relative weights of these domains vary according to the participant sample being compared. Amongst the participant sample as a whole, and amongst the males only, the items that load highest on Factor 1 are "Obeying the law", "Judges sending people to jail" and "telling the truth". In contrast, amongst the non-offenders, items that load highest on Factor 1 are "Keeping a promise to a stranger", "Saving the life of a friend" and "Keeping a promise to a friend", with the "truth, law and justice" issues loading on Factor 2. There is, however, one item loading that would appear to conflict with this interpretation of the factors. The item "How important is it not to take things that belong to other people?" loads on the "life and relationships" domain rather than on the "truth, law and justice domain".

2.58. Evaluation of "Importance" Ratings of SRM-SF Items: All Participants

Counts were made of the ratings given by participants of the "importance" of each of the items - i.e. whether the item was evaluated as "very important", "important" or not important". Previous studies that have carried out this task (e.g. Gregg, Gibbs & Basinger 1994; Palmer & Hollin 1998) have combined the "important" and "very important" counts into one category in order to be able to compare these ratings with the "not important" counts in a 2-by-2 Chi-Square analysis. In the current study, however, it was noted whilst tallying the importance ratings that a number of participants had chosen the "important" category as a way of expressing uncertainty about the importance of the item. For example, some participants had circled the "important" category but had then gone on to say "it's sometimes important" and sometimes not important". As in previous studies, the purpose of this exercise was to gain a measure of the levels of consensus amongst participants about the importance of the SRM-SF items – in order to establish that these items represent commonly held moral values. It was felt that only those responses where participants were unequivocal about the importance of the item should be categorised as representing consensus. Therefore, responses where participants expressed doubt were categorised as belonging to the "not important" category for the purposes of Chi Square analyses. The percentages of participants in each group rating items as unequivocally "important/very important" are presented in table 2.14 below:

Table 2.14: Percentage ratings of SRM-SF items as unequivocally "important" by participant group.

All Ages	Offenders (n=147)		Male Non-Offenders (n=149)		Female Non-Offenders (n=117)		Total (n=413)	
Item	n	%	n	%	n	%	n	%
Promise Friend	131	89.1%	147	98.7%	116	99.1%	394	95.4%
Promise Stranger	78	53.1%	119	79.9%	103	88.0%	300	72.6%
Promise Children	138	93.9%	142	95.3%	112	95.7%	392	94.9%
Tell Truth	104	70.7%	124	83.2%	106	90.6%	334	80.9%
Help Parents	129	87.8%	138	92.6%	110	94.0%	377	91.3%
Save Friend	144	98.0%	146	98.0%	112	95.7%	402	97.3%
Save Stranger	116	78.9%	121	81.2%	99	84.6%	336	81.4%
Stay Alive	94	63.9%	86	57.7%	74	63.2%	254	61.5%
Gibbs Steal	131	89.1%	146	98.0%	117	100.0%	394	95.4%
Obey Law	114	77.6%	129	86.6%	104	88.9%	347	84.0%
Judges Jail	87	59.2%	109	73.2%	96	82.1%	292	70.7%

A high proportion of participants rated most of the items unequivocally as "very important" or "important". However, the level of consensus for some items was lower than that reported in

previous studies (e.g. Gregg, Gibbs & Basinger, 1994, where a minimum of 88% from any participant group rated an item as important). There were some apparent differences between groups in the importance ratings of items. For most items, Female non-offenders had the highest percentage of participants rating items as unequivocally important, followed by male non-offenders, then male offenders. Chi Square analyses were carried out to compare between group differences in importance ratings for items. The results of the Chi Square analyses are presented in Table 2.15:

Table 2.15: Chi-Square statistics from comparisons of "importance" ratings of SRM-SF items by participant group:

Item	Offenders / Male Non-Offenders	Male non-offenders / Female non-offenders
Promising Friend	χ^2=11.80; df =1; p<0.01	χ^2=0.14; df =1; ns
Promising Stranger	χ^2 =23.88; df =1; p<0.001	χ^2=3.17; df =1; ns
Promising Child	χ^2= 0.29; df =1; ns	χ^2=0.27; df =1; ns
Telling Truth	χ^2 = 6.51; df =1; p<0.05	χ^2=3.05; df =1; ns
Helping Parents	χ^2 = 1.98; df =1; ns	χ^2=0.20; df =1; ns
Saving Friend	χ^2= 0.00; df =1; ns	χ^2=1.15; df =1; ns
Saving Stranger	χ^2= 0.25; df =1; ns	χ^2=0.47; df =1; ns
Staying Alive	χ^2= 1.20; df =1; ns	χ^2=0.84; df =1; ns
Not Stealing	χ^2= 9.69; df =1; p<0.01	χ^2=2.38; df =1; ns
Obeying Law	χ^2= 4.10; df =1; p<0.05	χ^2=0.32; df =1; ns
Judges Jailing	χ^2= 6.46; df =1; p<0.05	χ^2=2.94; df =1; ns

Results of Chi Square analyses comparing the ratings of importance of items of offenders and male non-offenders were statistically significant for the following items: "Keeping a promise to a friend", "Keeping a promise to a stranger", "Telling the truth", "Not taking things that belong to other people", "Obeying the law" and "Judges sending people to jail" (In all cases a higher proportion of non-offenders than offenders rating these items as "important"). None of the chi-square analyses comparing the importance ratings of male and female non-offenders were statistically significant.

2.6 DISCUSSION

The initial results comparing the participant groups correspond to the now established body of work that has identified differences between offenders and non-offenders on a variety of sociocognitive variables. Overall, offenders had significantly lower mean SRM-SF scores than male non – offenders. Furthermore, each of the SRM-SF items successfully discriminated between the offending and non-offending males. Finally, offender mean scores were particularly low for those SRM-SF items relating to offending behaviour and to the item "Telling the truth". On the face of it then, the preliminary findings would apparently lend support to evidence which suggests that offenders are demonstrating developmental delay in moral reasoning in relation to their non-offending peers. On

further investigation, however, some interesting findings emerged that raised questions regarding the interaction between offence status, gender, age and the variables under consideration.

When analysing the participant sample as a whole, no significant gender differences in SRM-SF mean scores were found among the non-offending sample. Comparison of individual SRM-SF sociomoral values and items across genders within the non-offender samples also failed to reveal significant gender differences favouring females to the extent that has been reported by some studies (e.g. Gregg et al, 1994; Palmer & Hollin 1998). There were significant differences favouring females on the "Affiliation" and "Life" values, and also on the individual item "Saving the life of a friend". However, these differences did not seem to reflect a consistent gender pattern across the questionnaire; males had higher scores than females on about half of the SRM-SF items, and had significantly higher scores on the item "Telling the truth".

Although no gender differences were found in SRM-SF scores taking the participant sample as a whole, once the participants were categorised according to age groups, significant gender differences did emerge in the youngest age group category, with females demonstrating significantly higher SRM-SF means than the non-offending males. Gender differences in the overall SRM-SF scores favouring females were no longer evident in either of the older 2 age group categories. This corresponds with those previous findings (e.g. Gibbs et al, 1992; Basinger et al, 1995; Ferguson & Cairns, 1996) that have found that gender differences favouring females on the SRM-SF are present only in early adolescence.

2.61 Problems Discriminating between Offender and Male Non-Offender Means in Age Group 1

In the youngest age group category, comparison of SRM-SF means failed to discriminate between the male offenders and non-offenders. Amongst the youngest males, there was little indication that the overall moral reasoning of offenders was "immature" in relation to that of non-offenders. By and large, the SRM-SF item means of offenders and non-offenders were very similar: both groups having item means that indicated a predominance of Kohlbergian Stage 2 reasoning. The one exception to this was to the SRM-SF item mean "Obeying the law", where non-offenders demonstrated significantly higher means than male non-offenders. Clear differences discriminating between male offenders and male non-offenders on the SRM-SF only emerged in the older age categories.

The sample sizes in the younger age – group categories were comparatively small, so results relating to these groups should be interpreted with caution. Nevertheless, the failure to find differences between male offenders and non-offenders in SRM-SF means in Age Group 1 is potentially interesting. It is of course possible that the lack of differences in means of offender males and non-offender males in the youngest age group category is idiosyncratic to the current study. Gregg et al, (1994) in their comparison of the means of offenders and non-offenders, attempted to obtain a "clean" non-delinquent sample by eliminating those students perceived by teachers as exhibiting conduct disorders. No conscious attempt was made to "clean" the non-offending sample in the current study. It may have been the case that amongst the younger "non-offending" males in this sample there were a higher proportion of those who "had not yet offended" or "had not yet been caught" relative to "true non-offenders" than in the older age groups. However, all of the non-offenders in the current study were in mainstream education at the time of the study and completed the SRM-SF in class as a voluntary exercise. A certain amount of "cleaning" of the non-offending sample would probably have occurred by default. Those pupils excluded or suspended from class for challenging behaviour at the time of the study were naturally eliminated from the sample – they were not present in school to carry out the task. It is also reasonable to assume that at least some persistent truants were absent from school at the time of the study. Furthermore, it is plausible that the youngest offenders in the current study may actually have been comparatively more "antisocial" in terms of persistent offending

behaviour than some of those in the older age group categories, bearing in mind that this group of adolescents had already been convicted by the courts at a very young age.

Categorising both non-offenders and offenders according to age groups may have raised issues that have not been highlighted in previous research using the SRM-SF. As was reported when presenting the overview to the measures, researchers comparing scores of offenders and non-offenders on the SRM-SF have often used participant samples with rather broad age ranges. The typical findings are that adolescent offenders seem to be "immature" in terms of moral reasoning in comparison with their non-offending peers. These findings are generally accompanied by evidence indicate that adolescent non-offenders are demonstrating at least some use of mature "Stage 3" reasoning, whereas the offenders seem to be exhibiting a predominance of "Stage 2" reasoning. However, it is presumably the inclusion of older non-offending adolescents in samples that enable this distinction to be made. It is hardly surprising that amongst a group of younger adolescent males that both non-offender and offender groups' moral reasoning scores indicate a predominance of Kohlbergian Stage 2 reasoning - this represents the norm for males in this age group, and previous work indicates that offenders' moral reasoning is best characterised by this Stage. On the other hand this then poses a problem when considering the utility of summary SRM-SF mean scores alone to discriminate between younger males. This issue is critical, especially when considering cognitive interventions with young people. Obviously, it is useful to be able to identify moral reasoning deficits amongst older offenders in order to point the way towards remedial interventions. Equally, it is apparent that older offenders typically have longer criminal histories than younger offenders and are more likely to be "well-established" within the criminal justice system. It stands to reason that interventions will stand a greater chance of success if the relevant cognitive risk factors are identified as early as possible, before a young person has become "entrenched" in offending behaviour. However, it seems from this sample at least that Global Moral Stage scores are not in themselves sufficiently sensitive for this purpose.

Separating out the offenders into distinct Age Groups may have raised a potential issue regarding the "global" nature of moral reasoning deficits amongst younger offenders, which wasn't picked up on when the offenders were all "lumped together" in one category. The general assumption made by

researchers who have identified poor performance amongst offenders on the SRM-SF seems to have been that this represents a "pervasive cognitive structural immaturity" (e.g. Gregg et al, 1994; Palmer & Hollin, 1998). However, on the basis of the findings from the current study, one would not be able to justify this claim in relation to the youngest offenders. As previously discussed, artefacts of the sample and lack of sensitivity of the measure may have created problems in this area. On the other hand it may be the case that younger offenders are not very different from non-offenders in terms of their overall moral reasoning, but are manifesting specific deficits that are in themselves risk factors in terms of global moral development. It was interesting to note that whilst younger offenders and male offenders seemed to be similar in SRM-SF scores across much of the questionnaire, the one area where non-offenders out-performed offenders was on the "Property and Law" value. Had one relied merely on summary SMR-SF mean scores amongst the younger adolescents, one would not have been able to identify this specific area of concern. This is of course the one domain that clearly differentiates between the two groups in behavioural terms.

An early disparity in "reasoning about the law" and involvement in criminal activity may have considerable knock-on effects resulting in global divergence in moral reasoning in later years. There are several reasons why this may be so. For example, persistent offending behaviour is likely to reinforce Stage 2 reasoning already present and is not conducive to a progression towards the "Mutual and Prosocial" reasoning typical of Stage 3. Criminal activity is more easily justifiable to the offender if he/she focuses on the pragmatic considerations of the situation rather than on the feelings of the victim(s) involved. There is a considerable body of research stretching back decades that demonstrates the existence of certain "techniques of neutralization" (Sykes & Matza 1957) and / or "cognitive distortions" (see e.g. Goldstein et al, 1998) that characterise the explanations given by offenders when seeking to rationalise or minimise their criminal behaviour.

Continued involvement in offending behaviour is likely to hinder the development of those "normal" social relationships considered essential in order to attain one's potential in terms of moral maturity (see e.g. Kohlberg, 1984, pp74-78). There is a substantial body of research demonstrates that offending behaviour is associated with membership of deviant peer groups (reviewed in e.g. Rutter et

al, 1998). Regardless of the controversy surrounding the potential causal link between "deviant peer-groups" and "offending behaviour", one thing seems to be clear: the more that young people engage in offending behaviour the more likely it is that they will be interacting with deviant peers; conversely it is less likely that they will be forming strong social ties with young people who are part of "mainstream" adolescent society. Persistent involvement in offending behaviour is also likely to be damaging to existing social relationships, for example by creating conflict within a family environment. Finally, it is clear that prolonged participation as a "client" of the criminal justice system is itself potentially inhibiting in terms of moral maturation. Kohlberg and colleagues noted at a relatively early stage that traditional prisons were unlikely to provide an environment that would facilitate mature sociolegal thought (Kohlberg et al, 1971; Scharf & Hickey, 1976). More recently, Taylor and Walker have provided empirical evidence that indicates that "moral climate" may serve as a mediating factor in terms of the relative success or failure of intervention programmes conducted in an institutional setting (Taylor & Walker, 1998).

As was pointed out earlier, the failure of the current study to identify global differences in SRM-SF scores between younger offenders and non-offenders may well have been idiosyncratic. However it is also possible that by treating adolescents of disparate ages as a homogenous group one runs the risk of distorting the picture regarding younger offenders' relative performance on the SRM-SF compared to their non-offending peers. This issue is one that needs clarification in future research using the SRM-SF.

2.62 Patterns of Change in SRM-SF Scores Observed across Age Groups

The data obtained from the comparisons of SRM-SF means across Age Groups are cross-sectional, so any inferences regarding developmental trends of the participant groups are of course speculative. Nevertheless, the comparisons across Age Groups did reveal some interesting findings that are worth discussion. As would be predicted on the basis of previous research, non-offender SRM-SF means increased across age group categories. For both male and female non-offenders the increase in scores between Age Group 1 and Age Group 3 was statistically significant – indicating developmental advancement in terms of moral reasoning. This apparent advancement was not reflected amongst the

offending sample. Offender mean SRM-SF scores showed little change across the age group categories. However, this finding should not be taken as an indication that moral reasoning of offenders remains fixed across adolescence. For one thing, it should be remembered that correlations between age and SRM-SF scores were statistically significant for all participant groups in the current study, including the offenders. This would seem to suggest that at least some developmental progression in moral reasoning takes place amongst offending adolescents. Furthermore, the offending group demonstrated considerable diversity in a number of important respects (e.g. in terms of types and location of settings represented and range of offences committed by participants) that may have had differential impacts on SRM-SF scores in the Age Group categories. Nevertheless, the findings do seem to suggest that developmental progression in moral reasoning that occurs amongst offenders is sluggish in relation to that of non-offenders.

Patterns of change amongst non-offenders across the age group categories were rather different for males than they were for females. Male non-offender SRM-SF means increased significantly across all age groups. The increase in male non-offender mean scores was particularly dramatic between age category 1 and age category 2 – by Age Group 2, male non-offender means were significantly higher than the offenders, and little different from those of the female non-offenders. The increase in scores for males non-offenders between Age categories 2 and 3 was also statistically significant. Whilst female scores also increased across the age categories, the pattern of increase was much steadier than for the males - significant differences were only found for the female scores between Age Group 1 and Age Group 3.

On the face of it the differences in male and female rates of progression would seem to correspond to a pubescent growth spurt occurring in the males. However, this may not be the entire explanation for the patterns of gender differences observed. The dramatic rise in male scores between Age Group 1 and Age Group 2 resulted in the males "catching up" with the females; however, for the males, the increase between the scores between Age Group 2 and Age Group 3 was also statistically significant. In contrast, the female scores seemed to level-off between Age Group 2 and Age Group 3. In fact, the increase in female non-offender scores and offender scores between these two categories was identical

in numerical terms (an increase from 262.93 to 272.41 for the female non-offenders, an increase from 216.26 to 226.87 for the offenders). Male non-offender means were higher than female means in Age Group 3. Whilst the difference was only slight (278.74 v 272.41), the standard deviation of the males was also higher (41.14 v 29.31). this might have been an indicator of differential patterns of "mature" stage use amongst the males and females. However, it should be noted that in Age Group 3, the samples were not well matched in terms of gender – there were much higher proportions of males (both offenders and non-offenders) in this category than females – and this may have resulted in a certain amount of distortion to the apparent age related trends.

The finding that the change in mean scores of the male non-offenders increased so dramatically across Age groups 1 and 2 is certainly one that should be investigated further, preferably with longitudinal data. The fact that it was between these particular two Age Groups that the dramatic increase in moral reasoning scores occurred is particularly interesting in terms of a UK context, as it raises an issue regarding the role of culture in stimulating moral reasoning. It is possible that the change in males may have been partially related to peculiarities of the UK educational system. In terms of school year groups, Age Groups 1 and 2 correspond to Year 9 (13-14 year olds) and Year 10 (14-15) respectively. The transition from Year 9 to Year 10 is a very important one for UK adolescents. At the end of Year 9, young people have to choose those specialist subjects that they are planning to take for their GCSE (16+) exams. In Year 10, young people's school-work starts to "count" towards grades in these final exams. Young people are suddenly confronted with the need to make decisions regarding their long term future.

2.63 Characteristics of the Scale:

Analysis of the inter-relationships of SRM-SF items revealed that all correlations were statistically significant, which may give some indication of the internal consistency of the scale. This was despite having subjected the SRM-SF to a blind "item by item" analysis and certain difficulties reported in scoring some items. As previously reported, some studies have tested the internal consistency of the SRM-SF by carrying out reliability analyses using Chronbach's alpha. Due to the high attrition rate on some items, it was felt that it would be inappropriate to carry out this procedure. The reasons for

this were as follows: In order to carry out a reliability analysis one would either have had to restrict the analysis to the relatively small (and not necessarily representative) subset of the data where every participant had provided scorable responses to every item, or one would have had to "replace the missing values" in some way. As the SRM-SF is a developmental measure, rather than a Likert scale, replacing missing values is problematic. The SRM-SF doesn't have the facility to include a mid-point "Not sure" category as many Likert scales do. If one replaced the missing values with an index of the item average, say the series median, one would potentially be distorting the picture regarding individual participants; if one replaced missing values with participant averages, one would potentially be distorting the picture regarding individual items.

An indication of the relative homogeneity/heterogeneity of the SRM-SF may, however, be arrived at through the results of the paired t-tests comparing SRM-SF sociomoral values and items and also from the results of the factor analyses. The combined results of the factor analyses and the inter-item paired t-tests are interesting, as they indicate that for this sample at least, the SRM-SF was not functioning in quite the same way for females as it was for males. Comparisons of the SRM-SF scores of the sociomoral values revealed that the questionnaire as a whole appeared to represent a more homogenous domain for the male non-offenders than for the other two groups. Paired t-tests found fewer significant differences between the values for the male non-offenders than for the offenders or female non-offenders. The results of the factor analyses raise issues regarding the "unitary" nature of the domain of the SRM-SF.

Results of the paired t-tests of the sociomoral values revealed that all groups tended to perform less well on the "Property and Law" and "Legal Justice" values than on the other values in the SRM-SF. This was particularly noticeable for the offenders and for the female non-offenders. The particularly low scores on "offence" related items amongst offenders is something that has been noted in previous research using the SRM-SF. For example, Gregg et al (1994) found that whilst "Non-delinquents" generally gave Stage 3 type reasoning responses to the area related to "Obeying the law", the "delinquent" sample generally employed Stage 2 reasoning. Gregg et al point to work by Gibbs et al (1984) who noted that although a Stage 2 reasoner may, in general, value law related norms, this

valuation is highly conditional. For example the young person may justify upholding the law by a reasoning structure such as "you might not get away with it". In the view of Gibbs et al (1984) adherence to the law is contingent in the respect that the young person may just as readily access instrumental considerations in order to justify law-breaking. On the other hand, the development of Stage 3 "Mutual and prosocial reasoning" acts as a "cognitive buffer" against anti-social behaviour. However, although the contingent valuation of Stage 2 reasoning may be more conducive to engaging in antisocial behaviour, it does not explain why - if "moral judgement" is assumed to represent a unitary domain - offenders should demonstrate poorer reasoning on "offence-related" issues than on other aspects of the domain. Gregg et al (1994) found evidence to indicate that a substantial minority of their predominantly "Stage 2" offending sample were applying "Stage 1 unilateral and physicalistic" reasoning to some of the offence related items in the questionnaire. They noted that there may have been a "situational artifact" accounting in part for their findings as the young offenders in their study had been recently convicted and were in the initial assessment phase of a custodial sentence (Gregg et al, 1994). Therefore there may have been an increased tendency to describe the consequences of law-breaking in terms of immediate and inevitable outcomes. However, Gregg et al (1994) go on to argue that this is unlikely to be the entire explanation, as poor performance on offence related items has also been identified amongst offenders who are not "incarcerated". They suggested that poorer performance amongst offenders on "law-related" items likely to be attributable in part to global deficits in moral reasoning (Gregg et al 1994). Unfortunately, they do not expand on this to explain why it should be the "law-related" items (rather than those concerning life and relationships) that act as particular identifiers of cognitive immaturity in moral reasoning.

It is of course the case that situational artefacts that may be relevant to "incarcerated" offenders' greater tendency to apply "Stage 1" reasoning to law related items also apply to those convicted offenders who are not in custody. Any young person who is sentenced by the courts is going to be made aware of salient punitive consequences of law-breaking. For example, when a non-custodial sentence is imposed on a young person the message "If you do this again you will be sent to jail" is one that comes across loud and clear from the Magistrates' bench. It seems plausible, then that

situational factors relating to experience of the criminal justice system may have had an impact on the performance of the young offenders in the current study on the "law related" values in the SRM-SF.

Whilst situational factors identified above may have accounted in part for offenders' poorer performance on the "law-related" values they fail to account for the comparatively poorer performance on these values that was demonstrated by the non-offenders, in particular by females. It seems likely that that due to the prevalence of youth crime, the immediate punitive consequences of law-breaking are made extremely salient to many adolescents by role models such as teachers and parents. One would, then, not be surprised if this impacted on SRM-SF scores even of non-offending adolescents. However, it is somewhat surprising that it was amongst the female-non offenders (rather than the males) that the greatest disparity in means between the law-related values and the other sociomoral values was observed.

It is possible that some clues to the differential performance of males and females on the "offence – related" values may be uncovered through examination of the relative variability demonstrated in SRM-SF scores by participant groups. Amongst all participants greater variability in SRM-SF scores tended to be found amongst those items relating to offending behaviour than to other SRM-SF items than to other items in the questionnaire. This greater variability may be an indication of minority "Stage 1" use amongst some participants due to situational factors already discussed. However, perusal of means and standard deviations of SRM-SF scores across all items reveals that in general, greater variability was evident in male non-offender scores than the other two participant groups. It was noted when discussing the findings from Age Groups that the scores and standard deviations found for males and females in Age Group 3 may have indicated differential patterns of mature Stage 3 and Stage 4 use. There may have been two confounding factors accounting for the females' apparent poorer performance on the offence related items. If it was the case that there was some evidence of minority Stage 1 use amongst all participants, but that a higher proportion of male non-offenders were making use of "Stage 4" reasoning relative to "Stage 3" reasoning than females, then this might account in part for the disparity between the males and the females. However, in order to

ascertain whether this was indeed the case, one would have to analyse the data in terms of frequencies of Stage use, rather than by reference to mean scores alone.

Inter-item comparisons within the socio-moral values raised some issues regarding the homogeneity of these so-called values. In particular, the inter-item comparisons amongst the "Contract and Truth" value called into question the validity of including the item "Telling the truth" as part of this domain, for adolescents at least. Participants tended to do less well on this item than on other items in the sociomoral value. This was particularly noticeable for the offenders and female non-offenders. This finding is hardly surprising when applied to the offender group – after all, a great deal of crime carried out by young people involves some form of deception. It would also not be terribly surprising if this finding applied to adolescents in general. After all, for many adolescents, the development of self-awareness and independence is likely to involve experimentation with avenues of experience that may conflict directly with parental wishes. Therefore, when it comes to telling the truth, for this age group, circumstances may frequently arise where pragmatic considerations ride high . However, the gender differences amongst the non-offenders relating to the "Truth" item – with females doing significantly less well on this item than males – are intriguing as they conflict with those previous studies that have compared SRM-SF item scores between male and female non-offenders (c.f. Gregg et al, 1994; Palmer & Hollin, 1998).

In a comparatively early review of the literature concerning gender differences in honesty, Burton (1976) pointed out that whilst a common stereotype of girls was that they were more obedient and more "trustworthy" than boys (Burton, 1976, p181) the empirical evidence regarding gender differences in honesty were conflicting. Burton reviewed several studies that indicated that in certain circumstances girls were more likely to cheat than boys (Burton, 1976). More recently, in a study carried out by Bear and Stewart (1990) investigating gender differences in perceptions of disciplinary interventions for common classroom transgressions, it was found that boys perceived interventions to be more acceptable for "cheating" than for "talking and getting out of one's seat". Girls on the other hand, tended to rate disciplinary interventions for both types of transgressions as being equally acceptable. Bear & Stewart (1990) suggested from these findings that girls were perhaps more likely

to view "talking and getting out of one's seat" as being "significantly more wrong" than boys. Another possible explanation of course is that boys view "deception" as being comparatively "more wrong" than girls. Some indications that this might be the case may be found in the "bullying" literature. Bullying in schools is widespread amongst males and females; however, it is well-documented that the nature of bullying tends to vary according to gender. Boys who bully are more likely to engage in overt physical bullying. Girls who bully are more likely to engage in indirect "relational aggression" using tactics of rumour, gossip and lies in order to socially exclude their victim (see e.g. Crick et al, 1999).

The results of the factor analyses differed from the previous finding obtained by Basinger et al (1995), who found when validating the measure that the factor structure of the SRM-SF was unidimensional. In the current study, Alpha factoring using the whole participant sample identified two factors with Eigenvalues over 1.00; this remained the case in Factor analyses carried out on the non-offending participant sample only, and also on the male participant sample only. Interpretation of the factors revealed that "life and relationship" issues had high loadings on one of the factors, and "truth, law and justice" issues had high loadings on the other. Whilst the same two factors emerged in all three factor analyses, their weightings were different according to the samples being compared. Amongst the participant sample as a whole, and amongst males only, scores on "truth, law and justice" issues accounted for much of the variance in the sample – loading high on factor one. This was perhaps predictable, as these were the issues where offenders' performance was particularly poor compared to their non-offending peers. However, amongst the male and female non-offenders, it was the "life and relationships" domain that loaded high on factor one and that accounted for much of the variance in the sample. Conversely, "truth, law and justice" issues were ones that loaded on the secondary factor. It was interesting to note that whilst few significant differences were observed on item means between the male and female non-offenders, the items that loaded high on the "life and relationship" factor were ones where females tended to do better than males; conversely, the items that loaded high on the "truth, law and justice" factor were ones where males tended to do better than females.

Relating the results of the factor analyses to previous research findings is problematic. Aside from the initial work carried out by Basinger et al (1995) when validating the SRM-SF, a review of the extant literature carried out by the author revealed only one other study (Krettenauer & Becker, 2001) where a factor analytic investigation of the SRM-SF had been undertaken. As in the current study, Krettenauer and Becker reported that they identified two factors with Eigenvalues over 1.00; however, their analysis was carried out on a shortened version of the measure (eliminating the items "Keeping a promise to a child", "Children helping their parents" and "Living even if one doesn't want to") and factor loadings for individual items were not reported. Neither of the aforementioned studies reported how the issue regarding the treatment of missing item responses was addressed. As noted earlier when discussing issues relating to the internal consistency of the measure, replacement of missing item responses (i.e. unscorable responses) for the purposes of a reliability analysis is problematic for a scale such as the SRM-SF. It was pointed out that if one replaced the missing values with an index of the item average, one would potentially be distorting the picture regarding individual participants; if one replaced missing values with participant averages, one would potentially be distorting the picture regarding individual items. Clearly this issue is also relevant when one is considering conducting a factor analysis on this type of data set. Whilst the author was unable to resolve this issue when she was considering carrying out a reliability analysis using the most recent version of SPSS, she was able to resolve the problem when it came to carrying out the Factor analyses by using the technique of "excluding missing values pair-wise" (this being an option offered by SPSS for Factor analyses, but **not** for reliability analyses). The studies by Basinger et al (1995) and Krettenauer and Becker (2001) present findings obtained from Factor analyses along with those obtained from reliability analyses; however, one cannot tell from these studies what "replacement of missing values" strategy was adopted for these sets of analyses. Without this information one is unable to draw direct parallels from the results of the factor analyses obtained in the current study, and previous research findings.

As far as the author is aware, there appear to have been only two studies that have presented findings from factor analyses carried out on all 11 SRM-SF items (Basinger et al, 1995, and the current study). Basinger et al, 1995 identified one factor only in their analysis; their findings would need to be replicated in order to validate the claim that the SRM-SF is unidimensional. The findings obtained in

the current study conflict with those of Basinger et al (1995) in that two factors were identified, with items relating to offending behaviour (or justice related issues) apparently loading on a different factor to the other items in the questionnaire. This issue is one that needs to be clarified in future research.

2.64 Importance Ratings of Items:

A high proportion of participants evaluated most items as being "important" or "very important". However, the percentages rating items as "important" were not as high as ones given in previous studies (e.g. Gregg et al 1994, Palmer & Hollin, 1998). Clearly, this finding almost certainly results from the way in which the "importance" ratings were coded in the current study – where particular attention was paid to ensure that only those ratings where participants had rated items as "unequivocally important" were assigned to the "important" category. However, it may be the case that as a result of coding the "importance ratings" in this way, a genuine difference between participants emerged that has not been identified in previous work. Female non-offenders had the highest percentages of participants rating items as important, followed by male non-offenders, then offenders. Differences between offenders and non-offenders were significant on a number of items, with a higher proportion of non-offenders rating items as important. Chi-Square analyses comparing male and female non-offenders were not statistically significant; nevertheless, magnitudes of the Chi-Square Statistic tended to be greater when comparing offenders and female non-offenders than they were when comparing offenders and male non-offenders. Barriga et al (2001) point out that it is unlikely that moral judgement per se is the only cognitive influence on moral behaviour. They see "moral self relevance" as being a potentially important factor. Future studies might want to pay more attention to this within the context of the SRM-SF. Clearly evaluation of the "importance" of a moral issue is not a direct or accurate indicator of personally relevant that issue is to an individual. However, modifications to the wording of the evaluative part of the SRM-SF item in such a way that does make it more personally relevant could be a way of shedding more light on this issue.

2.7 CONCLUSION

The initial analyses carried out on the participant sample as a whole apparently corroborated previous results using this measure amongst a similar population sample. However, on further investigation - with reference to specific Age Group categories- some conflicting findings emerged which are worthy of further exploration. Two main areas of concern emerged from the analyses presented in this chapter. Firstly, on the basis of the evidence from this sample, the discriminatory power of SRM-SF summary mean scores to identify differences in moral reasoning amongst younger male offenders and non-offenders was called into question. Secondly, issues were raised regarding the efficacy of overall SRM-SF scores to identify potential gender differences within this type of sample.

Gibbs et al (1992) consider Stage 3, Transition 3-4 and Stage 4 justifications as all representing examples of so-called "mature" reasoning. However, when comparing means between groups, Stages 3, Transition 3-4 and Stage 4 are treated as ascending points along a continuous scale.

By comparing SRM-SF scores amongst non-offending adults, it might be considered to be relatively straightforward to identify potential gender differences in Stage use with reference to summary mean scores alone – if one assumed that the majority of adults were using "mature" moral stages. One would expect that gender differences in Stage 3 versus Stage 4 reasoning corresponding to the Gilligan 1982 hypothesis would be reflected in significantly higher scores amongst males than females. However, even amongst adults, reliance on a summary mean score is not necessarily an accurate indicator of actual Stage use. It is of course hypothetically possible for someone to have a summary mean SRM-SF score that indicates a predominance of Stage 3 reasoning, without having given a Stage 3 response to any item in the questionnaire (e.g. by having an equal distribution of "Stage 2" and "Stage 4" responses). Amongst adolescents, where intuitively one would predict less consistency in "Stage use" than amongst children or adults, this may pose a particular problem. Reference to Standard Deviations of mean scores may give one additional clues to the variation of Stage use within samples; however, these are merely clues and are hardly satisfactory as a means of uncovering accurate patterns of Stage use.

Reliance on summary SRM-SF means to compare Stage use between offending and non-offending adolescent samples of disparate ages would seem to be particularly suspect. Use of a summary mean score amongst this type of population will say little about age related trends within the sample. As was pointed out in the overview to the SRM-SF, whereas some researchers have sought to identify age related trends amongst non-offending adolescents (e.g. Basinger et al, 1995), comparatively little attention has been paid to the identification of age-related trends amongst offending adolescents. As far back as 1990, Nelson et al highlighted the need for longitudinal studies to shed light on the question of whether the moral reasoning of offenders is "fixated" or whether it progresses at slower rates than that of non-offenders. Clearly, longitudinal studies may represent the "ideal-case" scenario in terms of establishing developmental trends amongst young people; however, setting up and maintaining longitudinal studies presents considerable practical difficulties that may not be easily resolved by the researcher – for example in terms of attrition rates during the course of the study (see e.g. Rutter, et al 1998:282-285). Not surprisingly, as in the current study, the bulk of the research that has sought to provide insight regarding "cognitive-deficit" versus "developmental-lag" in young offenders' reasoning has relied on cross-sectional data. However, it is remarkable to note that in a number of studies researchers have apparently failed to notice that the age-range of their offending sample may be a potentially relevant – and confounding - factor when seeking to shed light on this issue (e.g. Gregg et al, 1994; Palmer & Hollin, 1998). The results obtained from comparing the males in the current study illustrate this problem: – the apparent "immaturity" of offenders being restricted to the older Age Group categories. As was pointed out when presenting the results, sample sizes in the Age Groups may have had an impact on the findings in the current study. However, potential age confounds amongst offending samples would seem to be an area requiring further investigation in studies using the SRM-SF.

It became apparent during the course of the analyses that reliance on mean scores alone may have been inadequate in identifying potential gender differences in patterns of "mature" moral stage use. Regardless of the controversy surrounding the Gilligan (1982) hypothesis, differences in "mature" stage use within a non-offending adolescent sample may present one with a confound that has been neglected in the previous research. For example, if, for one reason or another, there are differences

within a non-offending adolescent sample in types of "mature" stage use, it will be difficult to establish an accurate measure of the relative maturity-immaturity of non-offenders to offenders. One could see how this could be particularly problematic if patterns of moral judgement of male and female non-offenders conform to the Gilligan (1982) hypothesis. Differences in Stage 3 / Stage 4 usage would result in inflated means (in terms of moral reasoning maturity) for males compared to females and also for male non-offenders compared to male offenders. This, clearly, is an issue that requires further investigation.

As a result of analyses carried in Chapter 2, it became clear that there were issues surrounding the SRM-SF requiring attention that were not adequately resolved by treating the measure as a continuous scale. In order to clarify some of these issues, it was felt necessary to carry out further statistical investigations of the SRM-SF, treating the data in categorical terms. Findings from these analyses are presented in Chapter 3, which follows.

CHAPTER 3: SUBSEQUENT ANALYSES OF SRM-SF DATA USING ALTERNATIVES TO SUMMARY MEAN SCORES

3.1 Overview

It will be recalled that in the discussion at the end of Chapter 2, issues were raised about the utility of relying on summary mean scores as the sole source of reference for information about patterns of moral reasoning between and within participant groups. The further analyses and findings presented in this chapter attempt to address some of those issues. The first part of this chapter reviews various alternative methods that researchers have commonly used in order to analyse moral reasoning data. Some potential advantages and disadvantages of these methods are described, and issues relating to the interpretability of some of the associated findings are presented.

The second part of this Chapter presents the results from the additional analyses that were carried out on the SRM-SF data set. Brief overviews of the rationale for and the aims of each of the additional analyses are given before presenting the results of the findings. The results are then discussed with reference to the findings obtained in Chapter 2 and also with reference to previous research.

3.11 Issues regarding the use of a "Global Stage Status" Scale as a summary index:

In Chapter 2, the bulk of the analyses were carried out with reference to SRM-SF mean scores. Gibbs et al (1992) consider the summary mean over all SRM-SF items to be the primary index for assessment of the SRM-SF. However, as pointed out in the overview to the measure in the introduction to this thesis, Gibbs et al (1992) provide an additional index to summary mean scores in the form of a "Global Moral Stage Status" scale (GMSS). The GMSS allows one to map the summary mean scores onto differential "developmental levels" of moral reasoning. As one finds that researchers using the SRM-SF typically relate findings obtained from summary means to "Global Moral Stages" it is worth exploring the properties of the GMSS in some detail.

The GMSS is a ten-point scale that represents the "developmental vicinity" in which an SRM-SF mean score may be located (Gibbs, Basinger & Fuller 1992). Table 3:1 shows the relationship between SRM-SF protocol scores and GMSS categories.

Table 3:1. Relationship between SRM-SF scores and GMSS categories:

SRM-SF Protocol Ratings	Corresponding Numerical Score	GMSS Categories	Summary Mean Point Boundaries
"Stage 1"	100	Stage 1	100-125
		Transition 1 (2)	126-149
"Transition 1-2"	150	Transition 2 (1)	150-174
"Stage 2"	200	Stage 2	175-225
		Transition 2 (3)	226-249
"Transition 2-3"	250	Transition 3 (2)	250-274
"Stage 3"	300	Stage 3	275-325
		Transition 3 (4)	326-349
"Transition 3-4"	350	Transition 4 (3)	350-374
"Stage 4"	400	Stage 4	375-400

Note: Whereas the SRM-SF protocol rating schema only has one "Transitional" category between each major stage, the GMSS incorporates two transitional levels between each major stage. The application of the term "Transition" to describe intermediate levels between major stages in both typologies is perhaps unfortunate and requires further clarification. In the GMSS, "transitional levels" are named by the major stage first with the "minor" stage represented in parentheses (Gibbs et al, 1992:40). Thus, a GMSS rating of "Transition 3 (2)" is deemed to represent majority use of Stage 3 reasoning with minority use of Stage 2 reasoning.

From Table 3:1 it can be seen that the range of SRM-SF scores encompassed by the GMSS categories are not the same for all developmental levels. The "Stage 2" and "Stage 3" category boundaries have twice as large a range as the other GMSS levels. Broadly speaking, aside from "Stage 2" and "Stage 3", the point boundaries of each GMSS category correspond to quartiles between the major Stages, each category having a range of approximately 25 points. For example, the point boundaries of the category "Transition 2(1)" span the range of SRM-SF scores from 150-174. In terms of proportions of pure or major Stage use, this would correspond to a Stage "mix" ranging from 50% Stage 1 use and 50% Stage 2 use to 26% Stage 1 use and 74% Stage 2 use. The "Stage 2" category spans the range of SRM-SF scores from 175-225, and the "Stage 3" category spans the range of SRM-SF scores from

275-325. Using the "Stage 3" category as an example, this would correspond to a pure stage mix ranging from 75% Stage 3 and 25% Stage 2, to 75% Stage 3 to 25% Stage 4.

It is also worth noting when considering the point boundaries of the GMSS the way in which scores corresponding to Transitional Stages in the SRM-SF typology (i.e. Transition 1-2, Transition 2-3 and Transition 3-4) have been assigned to GMSS categories. In each case, scores corresponding to a Transitional Stage in the SRM-SF typology are assigned to the higher of the two GMSS Transitional levels between each major stage. So, for example, an individual who gives Transition 2-3 responses to every item in the SRM-SF would be assigned a mean score of 250, which would correspond to a GMSS level of Transition 3(2). By implication then, a "Transitional Stage" moral justification in response to an SRM-SF item is more representative of the higher major Stage than the lower – rather than representing an exact mid-point between the two stages.

On the face of it the GMSS would seem to be a useful tool when seeking to explore differential styles of moral reasoning. However, on further examination it appears that representing patterns of moral stage use in terms of Global Stage Status may be problematic for a number of reasons.

Essentially, the utility of a "Global Stage Status Scale" as an additional summary index depends on its potential in terms of representing proportions of "major" and "minor" stage usage. Gibbs et al (1992) rationalised their use of GMSS point boundaries for the SRM-SF as "replicating" those used in their earlier Sociomoral Measure. In an article describing the construction and validation of the original SRM Gibbs et al (1982) justified their choice of point boundaries as follows:

"This boundary pattern was selected because it seemed to optimize the construction sample age trend" (Gibbs et al, 1982:900)

However, application of GMSS categories used in the original SRM to represent proportions of stage use from SRM-SF data may be inappropriate. In the manual for the SRM-SF, the original SRM from which the SRM-SF is described as consisting of "15 items and two moral dilemmas" (Gibbs, et al, 1992:38). But if one refers back to Gibbs et al (1982) one finds that:

*"Overall protocol ratings represent the mean stage level of subjects' scores on **eight** sociomoral norms"* (Gibbs et al, 1982: 900, emphasis added). However, overall SRM-SF scores represent averages over **eleven** items. It will be recalled that (aside from the broader range GMSS categories of "Stage 2" and "Stage 3") the GMSS categories represent percentage quartiles of adjacent major stage use. Clearly, percentage quartiles are more meaningful when applied to "averages out of eight responses" than they are when applied to "averages out of eleven responses". For example, the combinations of either six "Stage 3" responses and two "Stage 2" responses or four "Stage 3" responses and four "Transition Stage 2-3 responses would result in a score of 275 if averaged over eight norms. This score corresponds to the lower boundary of the GMSS category "Stage 3" (275-325) which in turn is inclusive of those scores that correspond to at least 75% Stage 3 use. On the other hand, the closest one can get to this score using combinations of either Stage 3 and Stage 2 responses or Stage 3 and Transition Stage 2-3 responses over eleven items is 272.7 (eight Stage 3 responses and three Stage 2 responses or six Stage 3 responses and three Transition Stage 2-3 responses), which corresponds to a GMSS of Transition 2(3). The next score up from this attainable by using a combination of either "Stage 3" and "Stage 2" responses or "Transition Stage 2-3 and Stage 3 responses is 281.8 (combination of either nine Stage 3 responses and two Stage 2 responses or seven Stage 3 responses. The shift from an eight norm scale to an eleven item scale also potentially impacts in theoretical terms on the interpretation of Transitional ratings. For example, in an eight norm scale a score of 250, which represents the lower end of Transition 2(3), relates to either four Stage 2 and four Stage 3 responses or to eight Transition 2-3 responses. However, in an eleven item questionnaire, this score is unattainable using a pure combination of Stage 2 and Stage 3 responses, corresponding merely to eleven Transition 2-3 responses.

It follows from the discussion above that the number of valid responses individuals supply to the SRM-SF becomes relevant when considering Global Stage Status. All questionnaires resulting in at least seven out of eleven scorable responses are considered valid for analysis and are assigned overall scores and GMSS ratings on the basis of overall averages. However, the weightings of individual responses to items vary according to the total number of responses given – greater weight being

attached to each individual response when seven valid responses are given than when eleven valid responses are given.

Gibbs et al (1982) defined Global Stage Status as follows:

".... the global score does not reflect any stage that accounts for less than 25% of the total number of scoring points assigned to the subject" (Gibbs et al, 1982:899)

As was pointed out by Jadack et al (1995) the concept of Global Stage Status relies on the assumption that individuals demonstrate relative consistency in moral reasoning across questionnaire items. Global Stage Status does not allow for individuals that are demonstrating disparate levels of moral reasoning in their questionnaire responses. For example, two Stage 1 responses and six Stage 4 responses over eight items would result in a score of 325, which corresponds to GMSS "Stage 3". Over eleven items, two "Stage 1" responses and nine "Stage 4" responses results in a score of 345.5 corresponding to a GMSS of "Transition 3(4)". Both GMSS ratings indicate a predominance of Stage 3 reasoning; however, in neither case has a Stage 3 response been offered. The possible response pattern described above is not entirely implausible when one considers that the SRM-SF is designed to be a group-administerable "pen-to-paper" task and that - as in the current study – test administrators may not always be in a position to monitor or probe every individuals' moral justifications.

Aside from the issues discussed above, the broader ranges of the "Stage 2" and "Stage 3" GMSS category boundaries relative to the other GMSS levels in the SRM-SF are rather curious bearing in mind the potential problems they pose in terms of discriminating patterns of Stage use between participant groups. The range of scores encompassed by the "Stage 2" boundary (175-225) makes it impossible to discriminate between those who may be "consolidating" Stage 2 reasoning and those who may be demonstrating upward transition away from this Stage. The category lumps together individuals who "need not" have given any evidence of mature moral reasoning (scores from 175-200) with those who have "necessarily" started to show at least limited evidence of moral maturity (scores from 201-225). In a similar vein, the broad range of the "Stage 3" category boundary (275-325) does not allow one to discriminate between those who may be consolidating Stage 3 reasoning

and those who are demonstrating upward Transition towards Stage 4. Clearly this creates additional problems when seeking to explore potential gender differences in patterns of mature stage use.

It is clear that there are a number of factors one should consider when using GMSS ratings to characterise SRM-SF data. These factors include the relative variance around the summary SRM-SF mean, the number of scorable responses supplied to the questionnaire and the appropriateness of the point-boundaries of the various GMSS levels. GMSS ratings are commonly presented as reference points to accompany SRM-SF summary means in the published research; however, as far as the author is aware, none of the issues raised above have, to date, received any attention of note.

3.12 Use of Modal Stages to Profile Moral Reasoning amongst Participants

A summary index that Gibbs et al (1992) suggest as an alternative to the summary SRM-SF mean is the modal protocol score attained by an individual over the eleven SRM-SF items.

A clear advantage of the mode over the mean in theoretical terms is, of course, that it allows one to treat SRM-SF data in categorical terms rather than as points on a continuous scale. This enables one to address issues raised by Gilligan (1982) and others, that Stage 3 and Stage 4 are best thought of as different styles of reasoning rather than as occupying different positions in a hierarchical typology. Furthermore, by using the mode as a summary index one is ensuring that one is profiling participants' moral reasoning with reference to actual scores that individuals have attained, rather than with reference to a hypothetical average. However, one of the disadvantages in using the mode is that its power as a discriminator of between-group differences in patterns of moral reasoning is rather limited. One is restricted to comparisons of the most frequently occurring stage categories between groups. Clearly, the observation that two groups have the same or different modal moral stages says little about the actual proportions of stage use within and between groups. It follows from this that as with the mean, the utility of the mode as a summary representation of participants' moral reasoning styles relies in part on the assumption that participants are being relatively consistent in their patterns of stage use across questionnaires. For example, if a participant group has a modal response of Transition Stage 2-3 but is also demonstrating high variability across questionnaire items then it is difficult to make the claim that this group's moral reasoning is "at" Transition Stage 2-3. For one

thing, the higher the variability in patterns of Stage use, the higher the likelihood is that the majority of participants are using styles of moral reasoning that fall outside the modal stage.

3.13 Intra-Individual Variation in Patterns of Stage Use

It has already been noted that the utility of a summary index such as the mean or the mode to characterise the moral reasoning of a participant group may be compromised if there is high inter-individual variability in scores within that group. Equally, when considering use of a summary index to characterise individuals' moral stage use, patterns of intra-individual variation are relevant. Walker & Taylor (1991:330) point out that the "caricature" of the Kohlbergian model implies "a strong notion of structural wholeness combined with the notion of sudden radical stage transitions" However, it will be recalled from the introduction to this thesis that this "caricature" represents Kohlberg's conceptualisation of moral competence, not of moral performance. As Krebs et al (1991) note, no stage theory expects all individuals to base all their judgements on one Stage only. Furthermore, whilst Kohlberg made claims about structural universalities in moral reasoning, he himself acknowledged the potential influence of a variety of mediating factors when considering differences in moral reasoning. For example Nisan & Kohlberg (1982), in their longitudinal study of moral reasoning in Turkey presented evidence indicating variability in moral reasoning according to the type of dilemma being discussed. They found that younger participants from both their "village" and "city" samples tended to do less well on dilemmas relating to law and punishment issues than to other issues under consideration. Nisan and Kohlberg suggested that lack of familiarity with a formal legal system may have partially accounted for their findings. It is also worth noting that the "Just Community" approaches that Kohlberg and colleagues applied to their moral reasoning interventions in prisons (e.g. Kohlberg, et al, 1971; Kohlberg et al 1975; Scharf & Hickey, 1976) schools (e.g. Blatt & Kohlberg 1975) and the kibbutz (e.g. Snarey et al 1985a; Snarey et al, 1985 b; Snarey, 1987) relied on the assumption that an individual's moral development may be moderated by certain environmental, social and educational factors.

Walker (1989) carried out a longitudinal study designed to test both the Kohlberg and Gilligan models of moral reasoning. Walker found no evidence to support Gilligan's hypothesis that Kohlberg's

typology was biased against females in terms of Stage 3 versus Stage 4 levels of reasoning. However, Walker did find some evidence to indicate that moral orientation (or content of moral judgements within stages) may differ according to whether dilemmas involved "real-life" or "hypothetical" considerations. Walker found that hypothetical dilemmas elicited more of the so-called "rights" type ("normative and fairness") orientations and that real life dilemmas elicited more of the "response" type ("utilitarian and perfectionism"). Carpendale & Krebs (1992) compared responses of young men to typical Kohlbergian moral reasoning dilemmas and dilemmas relating to business decisions. Whilst Carpendale & Krebs (1992) found considerable homogeneity within participant scores on two Kohlbergian dilemmas, far less internal consistency was demonstrated in terms of participants' moral reasoning when comparing scores on the Kohlbergian dilemmas and on the "business" oriented dilemmas. Carpendale & Krebs (1992) also found that average moral reasoning scores on the "business" dilemmas were lower than on the Kohlbergian dilemmas.

In a questionnaire such as the SRM-SF where a variety of different issues are under consideration one might predict that participants would be less internally consistent than when responding to standard Kohlbergian dilemmas. Furthermore, in the current study, there was some evidence from the item analyses presented in Chapter 2 to indicate that the nature of the issue under consideration may have impacted on response styles to SRM-SF items. For example, results of the paired t-tests of the SRM-SF sociomoral values presented in Chapter 2 revealed that in general participants tended to perform less well on those items relating to offending behaviour. However, one finds when reviewing the literature that exploration of trends in intra-individual variation in moral stage use across the SRM-SF is an area that, so far, appears to have been overlooked by researchers. When patterns of variation in Stage use have been investigated, these have generally been carried out as inter-item analyses to test the internal consistency of the scale, and / or as test / re-test conditions to establish that participants demonstrate relative stability in response set over time (e.g. Ferguson et al, 1994; Basinger, et al, 1995; Krettenauer & Becker, 2001). Given that one of the basic tenets of the Kohlbergian model is that individuals are comparatively structurally consistent in patterns of moral reasoning, it is perhaps hardly surprising that researchers testing the SRM-SF have focused their attention on establishing

reliability in the measure, rather than on possible sources of variation within individuals; however, this is an issue that is worthy of further investigation.

Information about patterns of intra-individual variation in stage use across different types of questionnaire items is useful if one is seeking to explore situational factors that may impact on moral reasoning. It is also of relevance when one is attempting to gain insight into potential intrinsic mediators of moral reasoning. Thoma & Rest (1999) point out that whilst the bulk of the moral reasoning research has focused on a wide range of external environmental factors that may be predictive of global changes in moral reasoning, more recently greater attention has started to be paid to intra-individual variations in patterns of moral judgement that may be predictive of developmental change. Thoma and Rest (1999) draw from earlier work carried out by e.g. Walker & Taylor (1991) that has explored intra-individual variations in moral Stage use with reference to patterns of "consolidation" and "transition". Walker & Taylor (1991) note in their discussion of Stage transitions in moral reasoning that there is some difficulty in the Kohlbergian research in reconciling the concept of "Structure" – associated with consistency in moral reasoning and consolidation of an existing Stage – with "disequilibrium"- associated with inconsistency in moral reasoning and transition between Stages. Walker and Taylor suggested on the basis of findings from their longitudinal (1991) study that most individuals' progression in terms of moral development can be characterised as a cyclical pattern involving greater or lesser internal consistency in terms of Moral Stage use (Walker & Taylor 1991:336). This may be assessed with reference to intra-individual variability in stage use with reference to their modal stage. High internal consistency is indicated when there is little variability in intra-individual stage use across moral reasoning items; this coupled with a negative skew in terms of proportion of responses below the mode is seen as being indicative of "consolidation" of an individual's modal moral stage. Lesser internal consistency is reflected by high stage "mixture"; this combined with a positive skew in terms of a majority of "subdominant stage responses" being above the modal stage (Walker and Taylor, 1991; Thoma & Rest, 1999) is seen as being predictive of transition towards a higher stage of moral development. Thoma and Rest (1999) also suggest that participants' reliance on a Kohlbergian moral framework as a means of interpreting sociomoral situations may vary according to whether they are in a period of "consolidation" or "transition".

Thoma and Rest (1999) found evidence to indicate that participants reliance on a Kohlbergian framework seemed to be highest when participants were in periods of apparent consolidation rather than transition (Thoma & Rest, 1999). Bearing in mind the nature of the sample in the current study (adolescents of differing ages rather than adults of similar ages) one would anticipate that at least some individuals would be exhibiting developmental progression between one stage and the next and that this progression might be reflected in terms of intra-individual variation in Stage use across SRM-SF items.

Information about patterns of intra-individual variation in stage use is clearly useful when exploring the "situational" impacts on moral reasoning and / or when investigating developmental trends within a participant sample. However, such information may also be useful when seeking to gain insight into the impact of factors such as gender as mediators on moral reasoning Stage use. For example, in a study carried out by Wark and Krebs (1996) where the nature of the moral issue was manipulated, it was found that whilst both males and females demonstrated lack of consistency across different types of moral dilemmas, females overall tended to be more consistent in their use of moral stage.
More recently, Haviv & Leman (2002) investigated the relationship between the nature of the moral dilemma, gender, gender role and moral reasoning. Haviv & Leman found no significant effect for gender *per se* on moral reasoning stage; however, they did find differences in the strength of pull towards "justice" and "care" content focus according both to the nature of dilemma and according to apparent gender role. It would seem then, that investigation into potential between-group differences in levels of consistency across the SRM-SF is an avenue worthy of exploration.

As was pointed out in Chapter 2, a recent meta analysis carried out by Jaffee & Hyde (2000) comparing gender differences in moral reasoning found that overall there were only small differences between males and females in terms of Stage 3 and Stage 4 reasoning. However, they also noted that certain factors appeared to moderate the magnitude of gender differences. These included the age of the participants, characteristics of the moral reasoning measure and treatment of data. Jaffee and Hyde found that that gender differences increased from childhood to adolescence but decreased sharply thereafter. They also noted that "moderately large" gender differences were found for those

measures – such as the SRM-SF – that did not make use of a dilemma. Finally, gender differences were apparently greatest when moral reasoning data was treated as a "categorical instead of a continuous outcome" (Jaffee and Hyde, 2000:720). It would appear then, that treating SRM-SF data in categorical rather than continuous terms would give one the greatest opportunity to identify underlying gender differences in moral stage use. One of the main areas of concern raised as a result of the analyses presented in Chapter 2 was of course, related to the effectiveness of the SRM-SF as an overall discriminator of gender differences in stages of moral reasoning.

3.2 Aims of the Empirical Research Presented in this Chapter

It is clear that aside from use of summary means there are a variety of alternative methods available that may be useful when investigating patterns of moral reasoning amongst young people. However, it is also clear that there are several theoretical and methodological factors that one needs to bear in mind when considering use of these alternatives. The analyses presented in the next section attempt to address some of the issues that have been raised in the overview presented above. As with the analyses presented in Chapter 2, one of the overall aims of the following section is to explore differences in patterns of moral reasoning with reference to offence status. The general prediction relating to differences in offence status is that through use of analyses involving alternatives to the summary mean one should be able to uncover patterns of between-group differences that correspond to those presented in Chapter 2. That is, through use of these alternative techniques one will find evidence that offenders are comparatively "immature" in terms of patterns of moral reasoning in relationship to their non-offending peers.

Another of the main aims of the following section is to explore differences in patterns of moral reasoning with reference to gender. Here, the overall objective is to seek to shed light on the Gilligan / Kohlberg debate by comparing findings obtained using alternatives to the summary mean with those presented in Chapter 2. A key prediction is that by use of categorical techniques one should be able to uncover gender differences in patterns of mature stage use that were not identified when treating SRM-SF as points on a continuous scale. More specifically, one would expect to find evidence that

females are demonstrating greater use of Stage 3 reasoning than males, and that males are demonstrating greater use of Stage 4 reasoning than females.

3.2 RESULTS OF SUBSEQUENT ANALYSES USING ALTERNATIVES TO SUMMARY MEAN SCORES

3.21 Exploration of between and within group differences in moral type with reference to Global Moral Stage Status (GMSS) ratings:

From the analyses carried out in Chapter 2 it was found that the summary mean SRM-SF scores of each of the participant groups were as follows: Offenders, mean = 222.59 (n = 147, S.D. = 30.11); male non-offenders 264.36 (n=149, SD=44.79); female non-offenders, mean = 262.48 (n=117, SD =29.88). In terms of corresponding "Global Moral Stage Status" (GMSS) ratings the offender mean would correspond to GMSS "Stage 2" (point boundaries encompassing mean scores ranging from 175-225), and both the male non-offender and the female non-offender means would correspond to GMSS "Transition Stage 3(2)" (point boundaries encompassing mean scores ranging from 250-274). On the basis of the mean scores and GMSS ratings given above, it would appear that offenders as a group are showing a predominance of "Stage 2" reasoning, whereas non-offenders are showing greater evidence of use of "Stage 3" reasoning.

As was pointed out in the introduction to this chapter, the GMSS categories used in the SRM-SF replicate those used in an earlier measure where moral reasoning summary scores represented means over eight sociomoral norms. It was noted that it is easier to relate GMSS levels to numbers of item responses over an eight item scale than it is to an eleven item scale such as the SRM-SF. Nevertheless, if one were to represent the GMSS scores in terms of ratios of "Stage 2" to "Stage 3" use over eleven items the following response profiles would be achieved: For the offenders the response profile over eleven items would be best represented by either a combination of nine "Stage 2" responses and two "Stage 3" responses, or seven "Stage 2" responses and four "Transition Stage 2-3" responses (both combinations result in a score of 218.18); for the non-offenders the response profile over eleven items is best represented by either a combination of four "Stage 2" responses and seven "Stage 3" responses, or eight "Transition Stage 2-3" responses and three "Stage 3" responses.

The applicability of GMSS ratings to characterise a particular participant group relies on the assumption that there is relative consistency in moral reasoning within the group; it follows from this that if one is to make meaningful comparisons between participant groups in terms of GMSS ratings that samples should not be heterogeneous in terms of variance. However, in this data set, some variability in SRM-SF scores – as evidenced by Standard Deviations around the means – is present within groups. Furthermore, some heterogeneity in variance between groups is evident from differences in the magnitudes of the Standard Deviations. For example, whilst the male and female non-offender SRM-SF means are practically identical (264.36 for the males, 262.48 for the females) and correspond to the same GMSS rating – "Transition 3(2)"- the magnitudes of the Standard Deviations are different, with males having a somewhat higher Standard Deviation (44.79) than the females (29.88). For the female non-offenders the SRM-SF scores encompassed by one Standard Deviation either side of the mean range from approximately 232 to 292, or GMSS levels "Transition 2(3)" to "Stage 3". However, for the male non-offenders the SRM-SF scores encompassed by one Standard Deviation either side of the mean range from approximately 219 to 309, or GMSS levels "Stage 2" to "Stage 3". It may be questionable, then, to characterise both male and female non-offenders as reasoning predominantly at GMSS level "Transition 3(2).

Aside from the issues mentioned above, representation of GMSS scores in terms of typical response profiles over eleven items (as presented earlier) is meaningful only if the majority of participants have supplied valid responses to all eleven items. However, it will be recalled that all SRM-SF questionnaires resulting in "seven or more" scorable protocols are considered "valid" in terms of data analysis. In order to investigate this issue further, the proportions of scorable responses provided to "valid" SRM-SF questionnaires was tabulated. Table 3:2 overleaf shows the breakdown of numbers of "scorable" responses supplied to valid SRM-SF questionnaires by participant group.

Table 3:2 Proportion of "scorable" responses supplied to the SRM-SF by participant group

	Offenders		Male Non-Offenders		Female Non-Offenders		Total	
Responses	N	%	N	%	n	%	N	%
7/11	21	14.3%	35	23.5%	26	22.2%	82	19.9%
8/11	38	25.9%	25	19.5%	25	21.4%	92	22.3%
9/11	40	27.2%	45	30.2%	28	23.9%	113	27.4%
10/11	29	19.7%	24	16.1%	25	21.4%	78	18.9%
11/11	19	12.9%	16	10.7%	13	11.1%	48	11.6%

From Table 3:2 above it can be seen that relating the GMSS levels to proportions of adjacent "Stage mix" over eleven SRM-SF items is problematic. Less than 12% of the total sample gave scorable responses to every item in the SRM-SF. The modal number of scorable responses given to SRM-SF items was nine out of eleven items. However, as this response pattern by no means typifies the response pattern of the participant groups, (representing as it does less than 30% of the total sample) relating GMSS ratings to a "standardised average" of nine out eleven responses would be inappropriate.

On the basis of the evidence in this sample it seems that there may be considerable difficulties when attempting to "match" SRM-SF response patterns to GMSS categories. This calls into question the utility of the GMSS as a frame of reference for characterising the moral reasoning stages of those responding to the SRM-SF. Bearing in mind that the GMSS scale is a device that is supposed to enable one to locate SRM-SF scores within moral reasoning "developmental vicinities" (Gibbs, Basinger & Fuller, 1992: 55), practical difficulties in applying the GMSS to SRM-SF data raise a number of non-trivial theoretical issues (for example, by raising issues regarding the potential "arbitrariness" of GMSS categories). These issues are taken up in the discussion at the end of this chapter.

Referring back to the data in Table 3:2 it can be seen that proportions of those giving only seven scorable responses out of eleven items appear to be somewhat higher amongst non-offenders than offenders. It will be remembered that whilst many of the offenders who supplied SRM-SF data were interviewed on a one to one basis, all the non-offenders completed the SRM-SF as a written task in a group setting, where it was more difficult to monitor for missing responses. However, the results of

Chi-Square analyses comparing between group differences in proportions of responses supplied were not statistically significant either when considering the participant sample as a whole ($\chi^2 = 7.36$; df =8; ns) or when comparing participant groups individually (offenders v male non-offenders: $\chi^2 = 5.72$, df =4, ns; male non-offenders v female non-offenders $\chi^2 = 2.10$, df =4, ns; offenders v female non-offenders $\chi^2 = 3.39$, df =4, ns).

Spearman correlations were carried for the participant sample as a whole and for each of the participant groups to investigate the relationship between number of scorable responses given and SRM-SF scores and also to investigate the relationship between age and number of scorable responses given. Investigations of these relationships were primarily of interest to the researcher for the following reasons: Firstly, on the basis of the questionnaires supplied in this sample it appeared that those providing only seven out of eleven "scorable" responses may have been producing less in terms of overall verbal output than those supplying "scorable" responses to all eleven items (for example it was noted in Chapter 2 when describing attrition rate in the initial sample that a substantial proportion of the "unscorable" responses supplied had arisen through participants leaving the item blank, or giving answers such as "don't know" or "same as before"). As the SRM-SF is a "production" measure rather than a "recognition" measure, identifying a relationship between verbal output and overall performance is relevant – particularly amongst those who have completed the task as a written exercise – as it may indicate a task-oriented confound relating to literacy skills. Secondly, age had originally been implicated as one of the factors accounting for overall attrition rate in the proportion of valid SRM-SF questionnaires supplied. The SRM-SF was devised in part in order to provide a practical alternative to the MJI that was suitable for use with children and young adolescents: The initial validation of the SRM-SF indicated that in the US it was suitable for use with children as young as 10 years old (Gibbs et al (1992), with overall attrition rate amongst this age group actually being lower than amongst high school students (Gibbs et al, 1992:42.). However, in the current study it was noted when describing treatment of data in Chapter 2 that the mean age of those supplying valid SRM-SF questionnaires was higher by approximately six months than the mean age of the original participant sample. Identifying a relationship between number of scorable responses supplied and

SRM-SF scores amongst valid questionnaires might point the way towards a potential confound between the nature of the task and SRM-SF performance for the younger participants.

Results of the correlations revealed significant positive relationships between number of valid responses given and SRM-SF scores for the participant sample as a whole ($n=413$, Spearman's rho = 0.12, $p<0.05$), for the offenders ($n=147$, Spearman's rho = 0.12, $p<0.05$) and for the male non-offenders ($n=149$, Spearman's rho $r=0.27$; $p<0.01$). However, for the female non-offenders, no significant relationship was found between the number of valid responses given and SRM-SF scores. Thus it appeared that the overall "amount of verbal output" produced may have been more of a factor accounting for males' performance on the SRM-SF than it was for the female participants. Difficulties associated with interpreting these findings are taken up in the discussion at the end of this chapter. No significant correlations between chronological age and number of scorable responses given were found for any of the individual participant groups.

If one disregards those questions regarding the appropriateness of applying single GMSS ratings to represent the moral reasoning of participant groups, a tool such as the GMSS may still provide one with a useful framework in order to explore the characteristics of the sample. Even if nothing else, the "levels" given in the GMSS scale should allow one to categorise SRM-SF means in a relatively systematic way. However, it was noted earlier in this chapter that the broader point boundaries encompassed by the "Stage 2" and "Stage 3" levels were potentially problematic in this respect. Therefore, Table 3:3 below, which summarises the proportions of participants "at" various GMSS levels also categorises those at "Stage 2" or "Stage 3" according to one of two subdivisions of these levels. The "Stage 2" category (scores ranging from 175-225) is further subdivided into "Stage 2a" – scores ranging from 175-200, and "Stage 2b" – scores ranging from 201-225. The theoretical justification for selecting the point boundaries for these subdivisions is that these subdivisions allow one to discriminate between those who "need not" have evidenced reasoning above Stage 2 when responding to any item in the SRM-SF - those assigned to Category 2a, from those who necessarily have given at least one response to an SRM-SF item that is scored higher than "Stage 2" – those assigned to Category 2b. This is clearly a useful distinction to be able to make when considering

differences in patterns of "immature" and "mature" stage use amongst offending and non-offending adolescents. In this sample, the GMSS of the offenders is given as corresponding to "Stage 2". Nevertheless, the overall mean - 222.59 - coupled with a Standard Deviation of 30.11 suggests that substantial proportion of this group were providing at least some evidence of reasoning above "Stage 2", which in itself is suggestive of at least some potential for a maturational shift in moral development. In line with the subdivisions made for the "Stage 2" category, the "Stage 3" category (scores ranging from 275-325) is further subdivided into "Stage 3a"-scores ranging from 275-300 and "Stage 3b" – scores ranging from 301-325. This subdivision allows one to distinguish between those who "need not" have used reasoning above Stage 3 and those who necessarily have used reasoning above Stage 3 (by giving at least one response that is scored at either "Transition Stage 3-4" or Stage 4). Being able to make a distinction of this nature is clearly of relevance when seeking to explore potential gender differences in patterns of mature Stage use.

Table 3:3 Proportions of respondents "at" different GMSS levels according to participant group.

	Offenders		Male Non-Offenders		Female Non-Offenders	
GMSS levels	n	%	n	%	n	%
Stage 1 = 100 - 125	0	0.0	0	0.0	0	0.0
Transition 1(2) = 126 - 149	0	0.0	0	0.0	0	0.0
Transition 2(1) = 150 - 174	7	4.8	4	2.7	0	0.0
Stage 2 = 175 - 225	81	55.1	26	17.4	14	12.0
Stage 2 a: Scores 175-200	*29*	*19.7*	*13*	*8.7*	*3*	*2.6*
Stage 2 b: Scores 201-225	*52*	*35.4*	*13*	*8.7*	*11*	*9.4*
Transition 2(3) = 226 – 249	28	19.0	17	11.4	25	21.4
Transition 3(2) = 250 – 274	24	16.3	41	27.5	38	32.5
Stage 3 = 275 – 325	6	4.1	50	33.6	38	32.5
Stage 3a: Scores 275-300	*6*	*4.1*	*29*	*19.5*	*29*	*24.8*
Stage 3b: Scores 301-325	*0*	*0.0*	*21*	*14.1*	*9*	*7.7*
Transition 3(4) = 326 - 349	1	0.7	6	4.0	2	1.7
Transition 4(3) = 350 - 374	0	0.0	5	3.4	0	0.0
Stage 4 = 375 - 400	0	0.0	0	0.0	0	0.0

From Table 3:3 it can be seen that the non-offenders are apparently showing that far greater evidence of "mature" Stage 3 ("mutual and prosocial") reasoning than offenders, whereas offenders are showing greater evidence of "immature" Stage 2 ("exchanging and instrumental") reasoning than non-offenders. Greater proportions of offenders have mean scores corresponding to GMSS level

"Stage 2" than non-offenders. For the offenders, "Stage 2" is the modal GMSS level, with 55.1% of the sample having means corresponding to this stage, whereas only 17.4% of male non-offenders and 12.1% of female non-offenders have means corresponding to this stage. Conversely, greater proportions of non-offenders have means corresponding to GMSS levels "Transition 3(2)" or "Stage 3" than offenders. For the male non-offenders the modal GMSS level is "Stage 3" with 33.6% of the sample having means corresponding to this stage. For the female non-offenders the modal GMSS levels are "Transition 3(2)" and "Stage 3" with equal proportions of participants – 32.5% - having means corresponding to these stages. However, only 16.3% of the offending sample have means that correspond to "Transition 3(2)" and even fewer (4.1%) have means that correspond to "Stage 3".

The creation of subcategories for GMSS levels Stage 2" and "Stage 3" has provided some potentially useful additional information regarding the characteristics of the sample. For example, whereas the modal stage of the offenders is "Stage 2" the majority of this group – 52/81 or 64% - are assigned to the subcategory "Stage 2b" which encompasses mean scores ranging from 201-225. In fact, one finds when taking the offending sample as a whole that whilst the group mean corresponds to GMSS level "Stage 2", the majority of these participants (75%) actually have means above 200, indicating at least some evidence of moral reasoning above this stage. This proportion is, however, still somewhat lower than that of the non-offenders (in the male non-offending group 89% have mean scores above 200, in the female non-offending group 97% have mean scores above 200).

The creation of sub-categories "3a" and "3b" reveal some apparent gender differences not readily identifiable when comparing groups assigned to the larger GMSS "Stage 3" category. Although the proportions of male and females assigned to GMSS level "Stage 3" are practically identical, (33.6% and 32.5% respectively), somewhat higher proportions of males (21/50 or 42%) have mean scores above 300 than females (9/38 or 24%). If one takes the non-offending sample as a whole, one finds that whereas approximately 22% of males have mean scores in the range 301-400, only 9% of females fall into this category. This would seem to suggest that evidence of Stage 4 "Systemic and Standard" reasoning (the "justice" orientation in Gilligan's terminology) may be greater amongst males than females.

Chi-Square analyses were carried out to compare between-group differences in the proportions of participants at varying GMSS levels. Due to low frequencies of participants being assigned to upper and lower ends of the GMSS scale, and in order to fulfil "expected" cell value" requirements, categories were combined in order to form a contingency table for Chi-square analyses as shown in Table 3:4 below.

Table 3:4. Categories used for Chi-Square analyses to compare proportions of participants at varying moral reasoning "developmental levels".

		Offenders		Male Non-Offenders		Female Non-Offenders	
Moral Stages	Range of Scores	n	%	n	%	n	%
Stage 1 to Stage 2a	100 – 200	36	24.5	17	11.4	3	2.6
Stage 2b	201 – 225	52	35.4	13	8.7	11	9.4
Transition 2(3)	226 - 249	28	19.0	17	11.4	25	21.4
Transition 3(2)	250 - 274	24	16.3	41	27.5	38	32.5
Stage 3a	275 - 300	6	4.1	29	19.5	29	24.8
Stage 3b to Stage 4	301 - 400	1		32	21.5	11	9.4

Chi-Square analyses revealed statistically significant differences between offenders and male non-offenders ($\chi^2 = 81.57$, df=5, $p<0.001$). The differences between offenders and non-offenders are mainly attributable to higher proportions of offenders than non-offenders being assigned to categories corresponding to lower levels of moral reasoning and to higher proportions of non-offenders than offenders being assigned to categories corresponding to higher levels of moral reasoning. The results of the Chi-Square analyses comparing offenders and non-offenders are in line with results of the ANCOVAs comparing overall means that were presented in Chapter 2 (these revealed significant differences between offenders and male non-offenders with offenders having lower means than non-offenders).

Results of a Chi-Square analysis comparing differences between male and female non-offenders was also statistically significant ($\chi^2 = 18.28$, df = 5, $p<0.01$). This finding contrasts with the results of the ANCOVA presented in Chapter 2 which did not reveal any significant differences in the overall

means of the two groups. The differences established by the Chi-Square analysis are mainly attributable to higher proportions of males than females having scores in the 301-400 range, indicating that males may be making greater use of Stage 4 reasoning than females. However, it should be noted that differences between the two groups are also attributable in part to higher proportions of males than females having mean scores of 200 or below. It appears that when carrying out the ANCOVA presented in Chapter 2 that the higher variability in male scores and higher proportions of males being represented at the lower ends of the GMSS may have acted as confounds precluding the identification of gender differences in the use of Stage 3 and Stage 4 reasoning.

3: 22 Exploration of Between and Within Group Differences in Moral Reasoning Styles Using Categorical Techniques

The analyses carried out in Section 3:21 above provide some useful additional information regarding the characteristics of the sample that was not revealed when carrying out the comparisons of group means that were presented in Chapter 2 (for example by revealing potential gender differences in use of Stage 3 and Stage 4 reasoning). However, the indices used as the bases for these analyses have still been participants' summary mean scores over the eleven items of the SRM-SF. As pointed out earlier, use of mean scores as summary indices relies on the assumption that differences in moral reasoning scores are best thought of as different points on a hierarchy - rather than as discrete categories or styles of moral reasoning as posited by Gilligan (1982) and others. The analyses in this section attempt to address this issue by exploring the different types of moral reasoning justifications given by participants using categorical techniques.

3. 23 Between Group-Differences in Modal SRM-SF Stages:

Modal stage use across SRM-SF items was tabulated for each participant. An investigation of the relationship between age and modal stage was carried out, using Spearman correlations. The result of this analysis was statistically significant ($n=413$, $r=.135$, $p<0.01$). This finding is in line with that obtained when comparing the relationship between age and SRM-SF mean scores that was presented in Chapter 2. Table 3:5 below shows the proportion of respondents modally "at" varying SRM-SF stages. It should be noted that the stages presented in this table relate to the actual protocol ratings

given when scoring SRM-SF items (i.e. "Stage 1", "Transition 1-2", "Stage 2" "Transition 2-3", "Stage 3", "Transition 3-4" and "Stage 4") and not the GMSS developmental levels relating to mean scores used in Section 3:21. In order to see how protocol rating scores relate to GMSS levels please refer back to the description that was given on page 73 and Table 3:1.

Table 3:5 Modal SRM-SF Responses by Participant Group

		Offenders		Male Non-Offenders		Female Non-Offenders		Total	
Protocol Ratings		(n)	%	n	%	N	%	n	%
Stage 1	100	(2)	1.4	(5)	3.4	(0)	0.0	(7)	1.7
Transition Stage 1-2	150	(3)	2.0	(1)	0.7	(0)	0.0	(4)	1.0
Stage 2	200	(51)	34.7	(19)	12.8	(16)	13.7	(86)	20.8
Transition 2-3	250	**(58)**	**39.5**	(38)	25.5	**(42)**	**35.9**	**(138)**	**33.4**
Stage 3	300	(29)	19.7	**(47)**	**31.5**	(37)	31.6	(113)	27.4
Transition Stage 3-4	350	(4)	2.7	(31)	20.8	(22)	18.8	(57)	13.8
Stage 4	400	(0)	0.0	(8)	5.4	(0)	0.0	(8)	1.9

Note: where bimodality was found, higher of two modes was used.

From Table 3:5 above it can be seen that the majority of respondents have modal protocol ratings that correspond to stages ranging from "Stage 2" to "Transition Stage 3-4" inclusive. Very few participants have modal scores corresponding to "Stage 1", "Transition "1-2" or "Stage 4". The low frequencies of respondents represented at upper and lower ends of the SRM-SF create problems for using Table 3:5 as the basis for overall Chi-Square analyses in terms of fulfilling "expected cell" requirements. Clearly, the most practical way to overcome this problem is to "collapse" the categories at the upper and lower ends of the SRM-SF (as was done when comparing proportions of participants at varying GMSS levels). However, this then potentially raises a question regarding the "contamination" of a discrete "Stage 2 category by including information relating to discrete categories "Stage 1" and "Transition Stage 1-2" (and the "contamination" of Stage 3-4 by including data relating to "Stage 4). A way of maintaining the integrity of the categories and to fulfil the "observed" and "expected" frequency requirements of the Chi-square analyses would be to treat those respondents represented in the "Stage 1", "Transition Stage 1-2" and "Stage 4" as "outliers". However, this potentially creates another problem as one then loses information regarding proportions of participants having modal scores at the upper and lower ends of the scale. Neither method then, is ideal. However, it was considered that by carrying out Chi-Square analyses using two different

methods (the first set adopting the technique of "collapsing" categories, the second set adopting the technique of eliminating "outlying" data) and comparing the findings obtained from both methods one should be able to make plausible inferences regarding group differences.

Results of the Chi-Square analyses comparing offenders and non-offenders were similar for both methods ("collapsing categories" and "eliminating outliers") that were adopted. Between group differences between offenders and male non-offenders were statistically significant (when "collapsing categories χ^2=48.77, df =3, p<0.001; when "eliminating outliers" χ^2=43.74, df = 3, p<0.001). The differences between offenders and non-offenders are mainly attributable to higher proportions of offenders than non-offenders being modally "at" Stage 2, whereas higher proportions of non-offenders than offenders are modally "at" Stage 3 or Transition Stage 3-4. The position regarding Transition Stage 2-3 is unclear. Whilst somewhat higher proportions of offenders than male non-offenders are modally "at" this stage, the observed difference between the two groups is not as great as it is for the other Stage categories. Broadly speaking, though, the results of the Chi-Square analyses comparing offenders and non-offenders replicated the findings presented in Section 3:21 above which established significant differences in the proportions of offenders and non-offenders represented at varying GMSS levels. In contrast to the findings obtained in Section 3:21, the results of Chi-Square analyses comparing male and female non-offenders failed to reveal significant differences between the groups for either method of analysis that was adopted ("collapsing categories" χ^2=4.32, df =3, ns; "eliminating outliers" χ^2=1.90, df=3, ns).

3.24 Investigation of patterns of intra-individual variability in SRM-SF stage use.

As reported earlier in this chapter, there is a conflict in the moral reasoning literature associated with reconciling "structural consistency" in moral reasoning that is fundamental to the original Kohlbergian model (and a requirement if one is to make meaningful comparisons between groups), with the need to account for a process of "disequilibrium" or "mechanism of change" that characterises an individual's progression from one stage to the next (e.g. Walker and Taylor 1991b). Amongst an adolescent sample one would expect to find that at least some of the participants were exhibiting "disequilibrium" in stage use predictive of developmental change; one might also expect that an adolescent sample would be less stable in terms of structural consistency than an adult sample

(assuming that for the majority of the adolescents moral potential is yet to be achieved and that for the majority of adults moral potential has been reached). However, researchers using moral reasoning measures amongst adolescents have typically made gross comparisons between groups on the basis of summary scores obtained from the averages of individual responses to several questionnaire items. Little attention has been paid to examining patterns of intra-individual variation in questionnaire responses which clearly may have an impact on findings obtained from group analyses.

The analyses presented in this section attempt to address this issue by investigating patterns of intra-individual variability with reference to range of stages used by participants, frequencies of non-modal judgements provided by participants, patterns of "positive" and "negative" skew around modal stage and also with reference to the chronological age of participants.

3.24i Variability in range of moral stages used across SRM-SF items by participant group

Table 3:6 below shows proportions of variability in the ranges between highest and lowest protocol scores obtained across SRM-SF items by participant group. To elaborate further, the higher the range of SRM-SF scores, the greater the disparity is between highest and lowest moral stage response evidenced across questionnaire items.

Table 3:6 Variability in range of moral stage scores by participant group

	male offender		male non-offender		female non-offender	
range	Count	Col %	Count	Col %	Count	Col %
50.00	7	4.8%	2	1.3%	5	4.3%
100.00	27	18.4%	27	18.1%	30	25.6%
150.00	56	38.1%	52	34.9%	50	42.7%
200.00	48	32.7%	33	22.1%	22	18.8%
250.00	7	4.8%	23	15.4%	10	8.5%
300.00	2	1.4%	12	8.1%	0	0.0%

From Table 3:6 it can be seen that there is evidence of intra-individual variability in patterns of stage use across the SRM-SF. None of the participants are completely consistent in using the same stage across all SRM-SF items and very few (less than 5% in any group) have a range of scores restricted to adjacent levels (i.e. 50 points) of the scale. The modal range of scores is 150 points. This corresponds to a stage mix of one and a half major stages, or, as an example, use of the range of

stages from Stage 2 to Transition Stage 3-4. Intra-individual variability appears to be higher amongst male non-offenders than in the other two groups. Higher proportions (24%) of male non-offenders have a range of scores of 250 points or higher than either the female non-offenders (8.5%) or the offenders (5.3%).

Chi-Square analyses were carried out to compare between-group differences in range of scores. Results of the analyses revealed significant differences between offenders and male non-offenders (χ^2 = 21.37, df = 5, $p<0.01$) and also between male and female non-offenders (χ^2=17.20, df=5, $p<0.01$). The result of a Spearman correlation comparing the relationship between age and range of scores was not statistically significant (n=413, rho = .066, ns).

3.24ii Investigation of frequencies of non-modal judgments by participant group

Patterns of variability in stage use were also compared by examining frequencies of non-modal judgements by participants – that is by comparing between-group differences in proportions of judgements made to questionnaire items that fell outside an individual's modal stage. The mean number of non-modal judgements for each of the participant groups was as follows: for offenders mean (n=147, SD=1.18)= 4.90, for male non-offenders mean (n=149, SD=1.14) = 4.91 and for female non-offenders (n=117, SD= 1.26) = 4.68. The result of a one-way ANOVA comparing between group differences in mean number of "non-modal" judgements made by participants was not statistically significant (F (2,410) = 1.503, ns). The result of a Spearman correlation investigating the relationship between age and frequency of non-modal judgements was not significant (n=413, rho = -.049, ns). Finally, results of a Spearman correlation comparing the relationship between "range of scores" and "frequency of non-modal judgements" was found to be statistically significant (n=413, rho = .365, $p<0.001$) indicating that those demonstrating higher ranges in patterns of stage use were also making a greater number of non-modal judgements.

3.24iii Investigation of Patterns of Skew around Modal Stage

In order to explore patterns of intra-individual variability further, patterns of skew around modal stage were investigated by participant groups. For each modal stage, participants were classified as exhibiting "negative skew" if a higher proportion of their non-modal judgements fell below the mode, as exhibiting "positive skew" if a higher number of their non-modal judgements fell above the mode and as demonstrating "no skew" if equal proportions of non-modal judgements fell both above and below the mode. Patterns of skew surrounding modal stages are presented in Table 3:7 below.

Table 3:7 Patterns of "Skew " surrounding modal stages by participant group

		Offenders		Male Non-Offenders		Female Non-Offenders		Total	
Modal Stage	Pattern of Skew	n	%	n	%	N	%	N	%
Stage 1	Negative Skew	0	0.0	0	0.0	-	-	-	-
	No Skew	0	0.0	0	0.0	-	-	-	-
	Positive Skew	2	100.0	5	100.0	-	-	7	100.0
Stage 1-2	Negative skew	0	0.0	0	0.0	-	-	-	-
	No skew	0	0.0	0	0.0	-	-	-	-
	Positive skew	3	100.0	1	100.0	-	-	4	100.0
Stage 2	Negative Skew	14	27.5	2	10.5	1	6.3	17	19.8
	No Skew	9	17.6	4	21.1	1	6.3	14	16.3
	Positive Skew	28	54.9	13	68.4	14	87.5	55	64.0
Stage 2-3	Negative Skew	43	74.1	16	42.1	16	38.1	75	54.3
	No Skew	2	3.4	7	18.4	9	21.4	18	13.0
	Positive Skew	13	22.4	15	39.5	17	40.5	45	32.6
Stage 3	Negative Skew	29	100.0	36	76.6	32	86.5	97	85.8
	No Skew	-	-	1	2.1	3	8.1	4	3.5
	Positive Skew	-	-	10	21.3	2	5.4	12	10.6
Stage 3-4	Negative Skew	3	75.0	30	96.8	22	100.0	55	96.5
	No Skew	-	-	1	3.2	-	-	1	1.8
	Positive Skew	1	25.0	-	-	-	-	1	1.8
Stage 4	Negative Skew	-	-	8	100.0	-	-	8	100.0
	No skew	-	-	-	-	-	-	-	-
	Positive Skew	-	-	-	-	-	-	-	-
All Stages	Negative Skew	89	60.5	92	61.7	71	60.7	252	61.0
	No Skew	11	7.5	13	8.7	13	11.1	37	9.0
	Positive Skew	47	32.0	44	29.5	33	28.2	124	30.0

From Table 3:7 it can be seen from the overall totals that there are higher proportions of those demonstrating "negative skew" around the mode than those demonstrating "no skew" or "positive skew" in all participant groups. This finding may be partially attributable to the way bimodality was tackled in the current study – by using the highest mode – which would bias the sample towards

greater evidence of negative skew. It can also be seen from the overall totals that the differences between participant groups in proportions of those demonstrating positive, negative or no skew are negligible; indeed, the results of a Chi-Square analysis comparing these between groups differences was not statistically significant ($\chi^2 = 1.32$, df = 4, ns).

If one examines Table 3.7 one can also see that despite the overall tendency towards negative skew in the sample, the overall proportions of those demonstrating positive and negative skew vary according to modal stage. Proportions of those demonstrating positive skew apparently decrease as modal stage increases in level. For example, when one examines the total proportions of positive, negative and no skew for those modally at Stage 2, one sees that 64% of those modally at this stage are demonstrating positive skew, whereas for those modally at transition Stage 3-4 less than 2% of those modally at this stage are demonstrating positive skew. It is possible that artefacts of the SRM-SF scale may have impacted on proportions of positive and negative skew that are observed in the sample. Clearly if an individual is modally at Stage 4 (the top end of the scale) one cannot be characterised as demonstrating positive skew around the modal stage. Transition Stage 2-3 represents the mid-point of the SRM-SF scale, *a priori,* then, one would predict that there would be greater evidence of negative skew amongst those modally at Stages above Transition Stage 2-3 than those modally below Transition Stage 2-3. To investigate this potential confound further, Spearman correlations were performed to investigate the relationship between "modal stage" and frequencies of judgments falling below and above the mode. Results of the Spearman correlations revealed a significant positive relationship between modal stage and frequencies of judgements falling below the mode (n=413, rho= .634, $p<0.001$) and a significant negative relationship between modal stage and judgements falling above the mode (n=413, rho= -.617, $p<0.001$). Finally, investigations of the relationships between age and "judgements falling below the mode" and "judgements falling above the mode" using Spearman correlations were carried out. However, no significant relationships were revealed between age and either of the variables under consideration (for the comparison between age and judgements falling below the mode rho (413) =.009, ns; for the comparison between age and judgements falling above the mode rho (413) = -.035, ns).

According to the "consolidation and transition" model (see e.g. Walker & Taylor, 1991, Walker et al. 2001) the pattern of stage use that is considered to be indicative of disequilibrium is characterised by positive skew coupled with high stage mix. However, when comparing differences in the range of stage use and frequencies of non-modal judgements amongst those demonstrating positive and negative skew in this sample, there was no evidence of such a relationship. Indeed, if anything, the pattern that emerged was contra to theory. Those demonstrating negative skew had a slightly higher mean stage range (mean (n=252, SD = 54.47) =164.48) than those demonstrating positive skew (mean (n=124, SD=57.95) = 162.50). Those demonstrating negative skew also had a slightly higher mean number of judgements falling outside the mode (mean (n=252, SD=1.24) = 4.96) than those demonstrating positive skew (mean (n=124, SD= 1.12) =4.77).

The findings obtained from comparisons of the proportions of positive and negative skew with degree of stage-mix were apparently contra to theory, and there would seem to be problems in applying the consolidation and transition model to SRM-SF data. Potential theoretical and methodological issues that may help to account for these findings are raised in the Discussion at the end of this chapter.

3.25 Comparisons of between-group differences in frequencies of stage use across the SRM-SF

In the discussion at the end of Chapter 2, the view was expressed that use of summary indices such as SRM-SF mean scores to characterise the moral stage of participants may have the effect of masking heterogeneity in patterns of stage use. The analyses presented in this section address this issue by comparing between-group differences in frequencies of moral stage use across the SRM-SF as a whole and also across individual SRM-SF items.

3.25i Comparisons of between-group differences in overall patterns of stage use

As a preliminary step towards summarising SRM-SF data in order to make meaningful between-group comparisons in patterns of stage use, frequency tables were drawn up for each individual major and transitional stage. Counts were made of frequencies of Stage category use – i.e. how many times participants gave responses corresponding to a particular stage category across all eleven SRM-SF items. However, it was found whilst tallying frequencies of stage use that few participants could be

considered to be making consistent use of one stage only (this finding is, of course, hardly surprising when one considers the patterns of intra-individual variation that emerged from the analyses presented in the previous section). For example, even for Transition Stage 2-3 use – the modal stage for participants as a whole – a very small percentage (2.9%) of participants gave six or more responses of this type across the whole questionnaire. It was also found that far less use was made of Stage Categories at the upper and lower ends of the SRM-SF (i.e. Stage 1, Transition 1-2 and Stage 4) –both in terms of proportions of participants using these stages and in terms of "consistency" of use within participants - than the mid-range Stage Categories. For example, only 13% of the total participant sample made any use of Stage 4 reasoning. Of those 13%, the majority (38/54 participants) used this Stage once only. To see the preliminary frequency table that was originally drawn up, please refer to Appendix 3.

For the purposes of statistical analysis, raw frequency data relating to Stage use was re-tabulated and is presented in Table 3.8 overleaf. For the mid-range stages (i.e. Stage 2, Transition Stage 2-3, Stage 3 and Transition Stage 3-4) the following categories were created in order to compare proportions of Stage use: "Not used", "Used once", "Used twice", "Used three times" and "Used four times or more". Whilst these categories are clearly hardly ideal in terms of comparing "high" and "low" use of individual stages, they do allow one to make at least some discrimination between lesser and greater use of the mid-range stages. For "Stage 1", "Transition Stage 1-2" and "Stage 4" (where it was not feasible to discriminate between greater and lesser stage use) data was merely tabulated according to whether participants had "used" or "not used" each individual stage.

For each of the individual stages Chi-Square analyses were carried out to compare between-group differences in patterns of stage use. For all stages 2-by-2 analyses were carried out to compare between-group differences in those "using" or "not using" a particular stage. For the mid-range stages separate Chi-Square analyses were carried out to compare between-group differences in greater or lesser stage use (using the categories "Not used", "used once", "used twice" "used three times" and "used four times or more"). For the comparisons between offenders and male non-offenders, the main predictions were that significantly higher proportions of offenders would use "immature" Stages of

moral reasoning than non-offenders, and that significantly higher proportions of non-offenders would use the "mature" Stages of moral reasoning than non-offenders.

Statistical comparisons exploring between group differences in patterns of stage use amongst male and female non-offenders were carried out in the same way as for the offender and male non-offender groups. For the comparisons between male and female non-offenders, a key area of interest was to establish whether there was evidence to support the Gilligan's (1982) claim that the differences between Stage 3 and Stage 4 correspond to differences in parallel female and male "voices" of moral reasoning rather than a developmental progression from one stage to the next. In order to provide support for this hypothesis one would predict that a significantly higher proportion of males would make use of "Stage 4" than females. Conversely, one would predict that a significantly higher proportion of females would make use of "Stage 3" reasoning than males. Summary statistics presenting the results of the analyses are presented in Table 3.9 on page 102. Key findings of note are described on page 102 after the Tables 3.8 and 3.9.

Table 3:8 Summary frequency data used to compare between-group differences in patterns of stage use

		Used		Not Used		1x		2x		3x		4 x	
		(n)	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)	%
Stage 1	Offenders	**(72)**	**49.0**	(75)	51.0								
	MNs	**(51)**	**34.2**	(98)	65.8								
	FN	**(17)**	**15.5**	(100)	85.5								
	All	**(140)**	**33.9**	(273)	66.1								
Stage 1.5	Offenders	(87)	59.2	(60)	40.8								
	MNs	(57)	38.3	(92)	61.7								
	FN	(40)	34.2	(77)	65.8								
	All	(184)	54.6	(229)	55.4								
Stage 2	Offenders	**(138)**	**93.9**	(9)	6.1	(26)	17.7	(34)	23.1	(36)	24.5	(42)	28.6
	MNs	**(104)**	**69.8**	(45)	30.2	(38)	25.5	(31)	20.8	(23)	15.4	(12)	8.1
	FN	**(92)**	**78.6**	(25)	21.4	(39)	33.3	(26)	22.2	(16)	13.7	(11)	9.4
	All	**(334)**	**80.9**	(79)	19.1	(103)	24.9	(91)	22.0	(75)	18.2	(65)	15.7
Stage 2.5	Offenders	(137)	93.2	(10)	6.8	(22)	15.0	(37)	25.2	(35)	23.8	(43)	29.3
	MNs	(123)	82.6	(26)	17.4	(36)	24.2	(36)	24.2	(26)	17.4	(25)	16.8
	FN	(113)	96.6	(4)	3.4	(24)	20.5	(34)	29.1	(22)	18.8	(33)	28.2
	All	(373)	90.3	(40)	9.7	(82)	19.9	(107)	25.9	(83)	20.1	(101)	24.5
Stage 3	Offenders	**(109)**	**74.1**	(38)	25.9	(45)	30.6	(31)	21.1	(23)	15.6	(10)	6.8
	MNs	**(123)**	**82.6**	(26)	17.4	(30)	20.1	(28)	18.8	(35)	23.5	(30)	20.1
	FN	**(108)**	**92.3**	(9)	7.7	(14)	12.0	(38)	32.5	(31)	26.5	(25)	21.4
	All	**(340)**	**82.3**	(73)	17.7	(89)	21.5	(97)	23.5	(89)	21.5	(65)	15.7
Stage 3.5	Offenders	(20)	13.6	(127)	86.4	(12)	8.2	(6)	4.1	(0)	0.0	(2)	1.4
	MNs	(102)	68.5	(47)	31.5	(37)	24.8	(33)	22.1	(14)	9.4	(18)	12.1
	FN	(68)	58.1	(49)	41.9	(33)	28.2	(16)	13.7	(9)	7.7	(10)	8.5
	All	(190)	46.0	(223)	54.0	(82)	19.9	(55)	13.3	(23)	5.6	(30)	7.3
Stage 4	Offenders	**(4)**	**2.7**	(143)	97.3								
	MNs	**(42)**	**28.2**	(107)	71.8								
	FN	**(8)**	**6.8**	(109)	93.2								
	All	**(54)**	**13.1**	(359)	86.9								

Table 3:9 Results of Chi-Square analyses comparing stage use across participant groups

	Offenders / Male Non-Offenders	Male / Female Non-Offenders
Stage 1: "count"	Not computed	Not computed
Stage 1: "binary"	$\chi^2 = 6.63$; df = 1; $p<0.05$	$\chi^2 = 13.36$; df = 1; $p<0.001$
Stage 1-2: "count"	Not computed	Not computed
Stage 1-2: "binary"	$\chi^2 = 12.97$; df = 1; $p<0.001$	$\chi^2 = 0.47$; df = 1; ns
Stage 2: "count"	$\chi^2 = 45.91$; df = 4; $p<0.001$	$\chi^2 = 3.67$; df = 4; ns
Stage 2: "binary"	$\chi^2 = 28.77$; df = 1; $p<0.001$	$\chi^2 = 2.64$; df = 1; ns
Stage 2-3: "count"	$\chi^2 = 16.58$; df = 4; $p<0.01$	$\chi^2 = 16.42$; df = 4; $p<0.01$
Stage 2-3: "binary"	$\chi^2 = 7.85$; df = 1; $p<0.01$	$\chi^2 = 12.89$; df = 1; $p<0.001$
Stage 3: "count"	$\chi^2 = 17.87$; df = 4; $p<0.01$	$\chi^2 = 12.62$; df = 4; $p<0.01$
Stage 3: "binary"	$\chi^2 = 3.08$; df = 1; ns	$\chi^2 = 5.46$; df = 1; $p<0.05$
Stage 3-4: "count"	$\chi^2 = 75.86$; df = 4; $p<0.001$	$\chi^2 = 5.78$; df = 4; ns
Stage 3-4: "binary"	$\chi^2 = 91.89$; df = 1; $p<0.001$	$\chi^2 = 3.04$; df = 1; ns
Stage 4: "count"	Not computed	Not computed
Stage 4: "binary"	$\chi^2 = 36.56$; df = 1; $p<0.001$	$\chi^2 = 19.57$; df = 1; $p<0.001$

O = Offenders, MN = Male Non-Offenders, FN = Female Non-Offenders

Note: "binary" analyses are those that merely compared whether participants had used or not used a particular stage using 2-by-2 Chi-Square analyses; "count" analyses are those that compared greater or lesser stage use.

3.25ii Results of statistical comparisons of stage use between offenders and non-offenders.

From Table 3:9 it can be seen that the majority of analyses comparing between group differences in patterns of stage use amongst offenders and male non-offenders were statistically significant. Comparisons of between-group differences amongst offenders and male non-offenders were statistically significant for all analyses except for the "Stage 3" analysis comparing proportions of participants "using" or "not using" this stage. In general, the between-group differences may be attributed to higher proportions of offenders than non-offenders making greater use of immature stages of moral reasoning (i.e. Stages 1 to Transition Stage 2-3) and higher proportions of non-offenders than offenders making greater use of mature stages of moral reasoning (i.e. Stage 3, Transition Stage 3-4 and Stage 4). Broadly speaking, then, the results of the analyses comparing offenders and non-offenders corroborate previous findings and might be taken as providing further with to the claim that offenders are immature in terms of moral reasoning compared to their non-offending peers. However, on further inspection one finds that characterising offenders and non-offenders in terms of relative immaturity or maturity may be an over-simplistic way of representing

differences between these groups. For example, between-group comparisons in patterns of Stage 2 use were statistically significant with higher proportions of offenders making greater use of this stage. However, if one refers back to the frequency data presented in Table 3:9 one finds that the majority of all participants – including the non-offenders – are apparently evidencing a certain degree of "moral immaturity" through the spontaneous production of at least one Stage 2 response. Similarly, between-group differences in patterns of Stage 3 use were statistically significant, with higher proportions of non-offenders making greater use of this stage. However, by referring back to the frequency data one finds that the majority of participants – including the offenders – are apparently able to access moral maturity with 74% of the offending sample using Stage 3 at least once (indeed, between-group differences amongst offenders and male non-offenders were not statistically significant for the Stage 3 used / not used analysis).

3.25iii Results of statistical comparisons of stage use between male and female non-offenders

The results of Chi-Square analyses confirmed the predictions that there were gender differences in patterns of mature Stage use in accordance with the Gilligan (1982) hypothesis. These findings contrast with those obtained in Chapter 2 where no significant differences were found in the overall SRM-SF means of male and female non-offenders. Chi-Square analyses revealed that between group differences in patterns of Stage 3 and Stage 4 use were statistically significant. Higher proportions of females used Stage 3 reasoning than males and higher proportions of males used Stage 4 reasoning than females. In terms of greater or lesser proportions of Stage use, between group differences for Stage 3 were also statistically significant, with females demonstrating a tendency to make greater use of this stage than males.

In addition to the gender differences described above, between group differences were also found for Stage 1 and Stage 2-3 binary comparisons of those "using" or "not using" these stages. A significantly higher proportion of males used Stage 1 reasoning than females; conversely, a significantly higher proportion of females used Transition Stage 2-3 reasoning than males. Between

group differences in greater or lesser Stage 2-3 use were also statistically significant, with females making more consistent use of this stage than males.

3.25iv Comparisons of stage use across SRM-SF items between participant groups

Between-group differences in frequencies of moral stage use across SRM-SF items were tabulated, and a frequency table summarising these differences may be found in Appendix 4. Statistical comparisons using 2-by-2 Chi Square analyses were carried out to investigate differences in Stage use between participant groups for each of the SRM-SF items. Carrying out the analyses required a large number of separate Chi-Square tests. For ease of reference a summary table - Table 3:10 - has been drawn up which indicates where significant differences were found. Chi-Square statistics for each of the separate analyses may be found in Appendix 5. Key findings of note from the analyses are described after Table 3:10 which is presented overleaf.

Table 3:10 Summary of results obtained from Chi-Square analyses of between group differences in stage use across SRM-SF items.

		Moral Stages						
Items	Comparisons	1	1-2	2	2-3	3	3-4	4
Promise Friend	Offenders / Male Non-Offenders	O*		O**			N***	
	Male & Female Non-Offenders					♀*		
Promise Stranger	Offenders / Male Non-Offenders			O**			N***	N*
	Male & Female Non-Offenders							
Promise Child	Offenders / Male Non-Offenders	O**			O**	O*	N***	N**
	Male & Female Non-Offenders							♂**
Tell Truth	Offenders / Male Non-Offenders			O***		N**	N***	
	Male & Female Non-Offenders			♀**			♂*	♂*
Help Parents	Offenders / Male Non-Offenders			O**			N**	
	Male & Female Non-Offenders				♀**			
Save Friend	Offenders / Male Non-Offenders				O**	N*	N*	N*
	Male & Female Non-Offenders	♂**		♂*	♀*	♀*		♂*
Save Stranger	Offenders / Male Non-Offenders				O**	N**		
	Male & Female Non-Offenders							
Stay Alive	Offenders / Male Non-Offenders			O*			N***	
	Male & Female Non-Offenders					♀*		
Not Steal	Offenders / Male Non-Offenders		O**	O*		N***		
	Male & Female Non-Offenders	♂**			♀***			
Obey Law	Offenders / Male Non-Offenders	O*	O**	O**	O*	N***	N***	N***
	Male & Female Non-Offenders				♀**			♂**
Legal Justice	Offenders / Male Non-Offenders			O**			N***	N*
	Male & Female Non-Offenders			♀**			♂*	♂*

*p<0.05, **p<0.01, ***p<0.001

Note: For the offender and non offender comparisons, asterisks indicate statistically significant differences; the letter "O" indicates that offenders were making greater use of the corresponding stage, the letter "N" indicates that non-offenders were making greater use of the corresponding stage. For the male and female non-offender comparisons, asterisks indicate statistically significant differences; symbols "♂" and "♀" indicate which gender demonstrated greatest use of the corresponding stage.

3.25v Results of Comparisons of Stage Use across SRM-SF Items Between Offenders and Non-Offenders

From Table 3:10 it can be seen that significant differences were found between offenders and male non-offenders for each of the SRM-SF items. In the main, where significant differences were found in stage use, higher proportions of offenders were using immature stages and a higher proportion of male non-offenders were using mature stages. There is some evidence from Table 3:10 that some items were more effective in discriminating between offenders and male non-offenders than others. For example, comparisons of stage use for the item "Obeying the law" revealed significant differences between offenders and male non-offenders for all stages. On the other hand between group

3.26 Recoding SRM-SF Protocols to Create a Unitary "Mature Stage"

As was discussed earlier, regardless of which stance one adopts in the Gilligan/Kohlberg debate, there is potentially a methodological problem associated with treating the "Mature Stages" as points on a continuous scale. Whilst Gibbs et al (1992) consider Stage 3, Stage 3-4 and Stage 4 judgements all to be representative of moral maturity, in terms of data analysis these stages are non-equivalent. Clearly, if one is seeking to explore potential gender differences in patterns of moral reasoning, it is useful to be able to discriminate between these levels in numerical terms. However, if one is seeking to compare populations in terms of relative overall maturity, then use of these differential levels may be problematic. From the results presented above it appears that male non-offenders were demonstrating greater evidence of Stage 4 reasoning than either the offender or the female non-offenders. This may have had an impact on the overall summary SRM-SF means by artificially inflating the male non-offender means relative to the other two participant groups.

In order to address this issue data from all SRM-SF protocols were recoded in order to impose arithmetical homogeneity on the differing mature levels. All protocols that had been assigned SRM-SF scores of Stage 3, Stage 3-4 or Stage 4 were recoded as "Stage 3" only. Table 3:11 compares the mean scores obtained using this method with the original summary mean scores presented in Chapter 2. Exploratory ANCOVAs using age as a covariate were carried out to compare between group differences on the recoded SRM-SF data. The results of these ANCOVAs, and comparisons with the original findings presented Chapter 2 are presented in Table 3:12. Comparisons of findings obtained using the different coding schemes are presented below Table 3:12 on page 109.

comparisons for the items "Saving the life of a stranger" and "Living even if you don't want to" - the items that make up the "life" value – only revealed significant differences for two stages.

From Table 3:10 it can be seen that the greater use of immature stages amongst offenders was most pronounced when comparing between patterns of Stage 2 use. For the majority of SRM-SF items between group differences were statistically significant for this stage, (higher proportions of offenders using Stage 2 than male non-offenders). Comparisons of Stage 1 and Transition Stage 1-2 reasoning amongst offenders and non-offenders revealed fewer between group differences overall than those obtained for Stage 2. This finding is hardly surprising when one considers that in general far less use was made of Stage Categories at the upper and lower ends of the SRM-SF than the mid-range categories.

Comparisons of patterns of mature Stage use between offenders and male non-offenders revealed that between group differences were most pronounced for Stage 3 and Transition Stage 3-4. These stages discriminated between offenders and male non-offenders for the majority of items – with higher proportions of non-offenders exhibiting mature reasoning. An exception to this was found for the item "keeping a promise to children" where a significantly higher proportion of offenders demonstrated Stage 3 reasoning than male non-offenders. However, it should be noted that on this item, modal Stage use for the non-offenders was particularly high – being Transition 3-4 for the male non-offenders. Finally, comparisons of patterns of Stage 4 use revealed significant differences between offenders and male-non-offenders on five items, with significantly higher proportions of male non-offenders making use of this Stage.

3.25vi Statistical Comparisons Between Male and Female Non-Offenders of SRM-SF Item Stage Use

Gender differences were found for patterns of Stage use on the majority of SRM-SF items (all save "Keeping a promise to a stranger" and "Saving the life of a stranger"). These findings contrast with those obtained in Chapter 2 when comparing mean scores of SRM-SF items, where gender differences were found for two items only. As was found when comparing Stage use across the questionnaire as a whole, there was some evidence to indicate that males were making greater use of Stage 4 reasoning

and females were making greater use of Stage 3 reasoning. When comparing patterns of mature stage use, between group differences were most marked for Stage 4 with significantly higher proportions of males demonstrating Stage 4 reasoning on five of the eleven SRM-SF items. When comparing patterns of Stage 3 use, it was found that significantly higher proportions of females demonstrated Stage 3 use on three of the eleven SRM-SF items. Finally, comparisons of patterns of Transition Stage 3-4 use revealed significant differences for two items - "Telling the truth" and "Judges sending lawbreakers to jail" with higher proportions of males using this stage on these items.

Comparisons of patterns of immature Stage use across SRM-SF items also revealed some gender differences. When comparing patterns of Stage 1 use it was found that a significantly higher proportion of males used this stage for two items - "Saving the life of a friend" and "Not taking things that belong to others". Gender differences in Stage 2 use were found for three items: on two of these items "Telling the truth" and "Judges sending people to jail" significantly higher proportions of females used this stage, on the item "Saving the life of a friend" higher proportions of males used Stage 2. Comparisons of Transition Stage 2-3 use revealed significant differences for four items – "Helping parents", "Saving the life of a friend", "Not taking things that belong to others" and "Obeying the law". On all of these items, higher proportions of females than males used this stage.

The results of the chi-square analyses indicated that gender differences appeared to relate not only to overall stage use, but also to the nature of the issue under consideration. As described above, comparisons of Stage use across SRM-SF items revealed some differences in patterns of Stage use corresponding to the Gilligan (1982) claim that Stage 3 reasoning is more prevalent amongst females and Stage 4 reasoning is more prevalent amongst males. However if one refers to Table 4:10 it can be seen that there is some evidence to suggest that males were also exhibiting greater moral maturity in terms of SRM-SF stage use than females on the "Justice" related items in the questionnaire. For example, on the items "Telling the truth", "Obeying the law" and "Judges sending people to jail" gender differences that were found indicated that females were making more use of an immature stage (either Stage 2 or Transition 2-3) where males were making greater use of one or more of the mature stages (Transition 3-4 and/or Stage 4).

Table 3:11 Descriptive Statistics comparing "original" SRM-SF means with re-coded means

		Original SRM-SF	Upper Limit = Stage 3 SRM-SF
Participant Group	N	Mean (SD)	Mean (SD)
Offenders	147	222.59 (30.11)	220.92 (27.43)
Male Non-Offenders	149	264.36 (44.79)	250.60 (33.57)
Female Non-Offenders	117	262.48 (29.88)	254.86 (22.99)
Total	413	248.96 (40.97)	241.24 (32.43)

Table 3:12 Comparison of results obtained using ANCOVAs on "original" SRM-SF means and recoded means

	Offenders and Male Non-Offenders	Male and Female Non-Offenders
	F (df)	F (df)
Original SRM-SF ANCOVAs	111.49 (1,293)***	0.50 (1,263)
Upper Limit = Stage 3 SRM-SF ANCOVAs	86.94 (1,293)***	5.45 (1,263)*

From Table 3:12 it can be seen that recoding the data changed the outcome that emerged when comparing male and female non-offenders. In the original findings, whilst male non-offenders had slightly higher means than female non-offenders, between group differences were not statistically significant. However, once the data had been recoded, the position was reversed, with females having significantly higher mean scores than males.

3.26i Comparisons of Recoded Individual SRM-SF Items by Participant Group

Exploratory ANCOVAs (using age as a covariate) were carried out using the recoded data to compare between group differences in mean scores on each of the individual SRM-SF items. Summary statistics showing mean scores of items using the original and modified coding schemes and showing the results of ANCOVAs carried out on the original and recoded data are presented in Tables 3:13-3:16. Key findings of note from the ANCOVAs are described after Table 3:16 on page 111.

Table 3:13 Item means obtained using Gibbs, Basinger & Fuller (1992) standard scoring procedure

Original	Offenders			♂ Non-Offenders			♀ Non-Offenders			Total		
Values & Items	N	Mean	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
Promising Friend	*130*	*236.92*	*66.63*	*137*	*279.93*	*60.47*	*113*	*288.50*	*47.72*	*380*	*267.76*	*63.23*
Promising Stranger	*113*	*225.22*	*54.33*	*105*	*266.67*	*71.95*	*93*	*267.74*	*54.50*	*311*	*251.93*	*63.98*
Promising Children	*124*	*239.11*	*67.94*	*131*	*301.53*	*68.20*	*98*	*293.37*	*62.85*	*353*	*277.34*	*72.27*
Telling Truth	*133*	*218.80*	*45.89*	*124*	*266.13*	*66.43*	*99*	*237.37*	*56.85*	*356*	*240.45*	*60.11*
Help Parents	*129*	*234.50*	*55.83*	*130*	*268.85*	*60.77*	*103*	*269.90*	*43.35*	*362*	*256.91*	*56.87*
Saving Friend	*127*	*236.22*	*66.87*	*121*	*261.16*	*76.51*	*92*	*279.35*	*39.33*	*340*	*256.76*	*66.73*
Saving Stranger	*95*	*238.95*	*57.05*	*96*	*265.10*	*64.12*	*65*	*256.15*	*56.26*	*256*	*253.13*	*60.47*
Staying Alive	*113*	*241.15*	*50.10*	*98*	*265.31*	*66.35*	*90*	*273.33*	*52.04*	*301*	*258.64*	*57.95*
Stealing	*126*	*194.05*	*57.83*	*125*	*216.80*	*81.80*	*98*	*231.63*	*63.97*	*349*	*212.75*	*70.42*
Obeying Law	*116*	*194.83*	*63.38*	*116*	*282.33*	*89.26*	*90*	*257.78*	*70.68*	*322*	*243.94*	*84.48*
Judges Jailing	*104*	*195.67*	*58.75*	*115*	*249.57*	*86.47*	*86*	*228.49*	*59.18*	*305*	*225.25*	*74.02*

Table 3:14 Item means obtained when imposing an "upper limit" of Stage 3

Upper Limit = Stage 3	Offenders			♂ Non-Offenders			♀ Non-Offenders			Total		
Values & Items	N	Mean	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
Promising Friend	130	233.85	62.32	137	267.15	47.53	113	277.43	35.98	380	258.82	53.46
Promising Stranger	113	223.01	49.57	105	251.43	53.01	93	258.61	42.11	311	243.25	50.98
Promising Children	124	237.50	65.53	131	273.28	46.12	98	272.45	46.60	353	260.48	56.31
Telling Truth	133	218.05	43.67	124	252.42	48.50	99	233.33	50.00	356	234.27	49.27
Help Parents	129	231.79	51.13	130	257.31	46.22	103	263.11	32.05	362	249.86	46.56
Saving Friend	127	235.83	66.34	121	253.31	67.31	92	277.72	37.33	340	253.38	62.37
Saving Stranger	95	237.89	55.41	96	259.38	56.81	65	253.85	53.26	256	250.00	56.01
Staying Alive	113	239.82	47.79	98	253.06	51.18	90	265.00	40.68	301	251.66	47.93
Stealing	126	193.25	55.98	125	212.40	74.77	98	228.57	59.20	349	210.03	65.54
Obeying Law	116	191.81	55.93	116	255.17	63.38	90	247.78	58.02	322	230.28	65.86
Judges Jailing	104	193.27	52.63	115	231.74	63.98	86	223.26	49.53	305	216.23	58.69

Table 3:15 Results of inter-item ANCOVAs obtained from original item means

Original	Offender ♂ Non-Offender	♂ Non-Offender ♀ Non-Offender
Item	F (df)	F (df)
Promise Friend	*35.30 (1,264)* ***	*3.62 (1,247)*
Promise Stranger	*30.92 (1,215)* ***	*0.62 (1,195)*
Promise Children	*59.27 (1,252)* ***	*0.24 (1,226)*
Tell Truth	*48.11 (1,254)* ***	*8.32 (1,220)* **
Help Parents	*28.44 (1,256)* ***	*0.87 (1,230)*
Save Friend	*14.33 (1,245)* ***	*7.60 (1,210)* **
Save Stranger	*10.90 (1,188)* **	*0.20 (1,158)*
Stay Alive	*13.63 (1,208)* ***	*2.22 (1,185)*
Not Stealing	*8.25 (1,248)* **	*3.23 (1,220)*
Obey Law	*78.92 (1,229)* ***	*2.17 (1,203)*
Legal Justice	**31.72 (1,216)** ***	**1.76 (1,198)**

Table 3:16 Results of inter-item ANCOVAs using recoded SRM-SF data

Recoded SRM-SF data	Offender ♂ Non-Offender	♂ Non-Offender ♀ Non-Offender
Item	F (df)	F (df)
Promise Friend	*27.14 (1,264)* ***	*6.89 (1,247)* **
Promise Stranger	*22.09 (1,215)* ***	*2.76 (1,195)*
Promise Children	*28.61 (1,252)* ***	*0.04 (1,226)*
Tell Truth	*38.06 (1,254)* ***	*5.55 (1,220)* *
Help Parents	*22.59 (1,256)* ***	*3.45 (1,230)*
Save Friend	*9.29 (1,245)* **	*14.24 (1,210)* ***
Save Stranger	*8.52 (1,188)* **	*0.03 (1,158)*
Stay Alive	*6.29 (1,208)* *	*5.24 (1,185)* *
Not Stealing	*6.80 (1,248)* *	*4.29 (1,220)* *
Obey Law	*67.10 (1,229)* ***	*0.11 (1,203)*
Legal Justice	25.51 (1,216) ***	0.17 (1,198)

By comparing Tables 3:15 and 3:16 it can be seen that once the data were re-coded a greater number of significant differences between male and female non-offenders on SRM-SF items were revealed than were when using the original coding scheme. When using the original coding scheme, males and females had significantly different means on two items only: "Telling the truth" and "Saving the life of a friend" (with males having significantly higher means on the item "telling the truth" and females having significantly higher means on the item "Saving the life of a friend". Using the modified coding scheme additional differences were found between male and female non-offenders for the items "Keeping a promise to a friend", "living even if you don't want to" and "not taking things that belong to others" – in all cases females having significantly higher means than males.

3:3 DISCUSSION

The analyses presented in this chapter sought to address a number of theoretical and methodological concerns. Some of these concerns had emerged as a result of findings obtained in Chapter 2, other issues emerged when considering the relative merits of alternative techniques of analysis in Section 3:1. Furthermore, the results of some of the analyses raised questions that in themselves may be worthy of investigation in future research. The first part of this section presents a broad summary of the overall outcomes and issues to emerge from the main body of findings presented in this chapter. The second part of this section considers each individual set of analyses in more detail, discussing the key findings that emerged, and, where relevant, relating these findings to those presented in Chapter 2.

3:31 Overall outcomes:

One of the most striking impressions that emerges from the results of the analyses carried out in this chapter is that the technique of analysis that one adopts may have a critical impact on the findings that are obtained. This impression is particularly salient when one examines the results of the findings obtained with reference to gender differences in this chapter and compares them with those presented in Chapter 2. The majority of the findings relating to offence status were, broadly speaking, comparable to those presented in Chapter 2 (providing evidence to indicate that offenders were comparatively immature in terms of moral reasoning in relation to their non-offending peers). However, the findings relating to gender differences in this chapter were in contrast to those obtained when making comparisons with reference to SRM-SF summary means. Whereas in Chapter 2, comparisons of overall SRM-SF means revealed no significant differences between male and female non-offenders, gender differences materialised from a number of analyses that were presented in this chapter. Furthermore, whilst the overall picture regarding offenders' moral reasoning was, in general, indicative of lesser use of mature levels of moral reasoning when compared to non-offending peers, it became clear during the course of the various findings that were presented that this comparative immaturity was manifested in terms of degrees of performance rather than in terms of overall global competence.

3.32 General Observations Regarding Variability in the Sample

The point was made in the overview to this chapter that in order to accurately represent the moral reasoning patterns of participant groups one needs to pay careful attention to inter-individual and intra-individual variation in Stage use. It was clear in the current study that both inter-individual variability in SRM-SF means within groups and intra-individual variability in stage use across SRM-SF items were indeed factors that had acted as confounds when representing moral reasoning patterns in terms of SRM-SF summary means.

It is not entirely clear from the previous SRM-SF research whether the level of inter-individual variability in mean scores in the current study is a typical phenomenon. Firstly, not all studies using the SRM-SF have presented Standard Deviations alongside summary means (e.g. Krcmar & Valkenburg, 1999); secondly, the scoring method adopted in the current study (scoring questionnaires by item rather than by participant) increased the potential for introducing heterogeneity of intra-individual variance in Stage use. Finally, due to the age range of the participant groups one would predict a certain level of variability in mean scores as a result of differences in maturational development. However, from a perusal of the literature it seems that whilst the overall standard deviation of the non-offending sample is somewhat higher than that cited in some studies (e.g. Barriga et al who reported a standard deviation of 25.00 around their summary mean of 305) it is comparable to findings obtained in others (e.g. Palmer and Hollin, 1998 who reported a standard deviation of 41.26 around a summary mean of 285.57 for male non-offenders and a standard deviation of 35.19 around a summary mean of 304.06 for female non-offenders). Therefore, it would appear that despite the scoring method adopted, the overall level of variability observed in the current study is by no means aberrant.

Differences in the magnitude of variability were observed between participant groups, with male non-offenders having somewhat higher standard deviations than either the female non-offenders or the male offenders. As with the overall standard deviations it is difficult to tell whether this apparent gender difference in the magnitude of variability is typical - in some studies where gender comparisons have been made, the individual means for males and females have not been presented

(e.g. Palmer & Hollin, 1997; Palmer & Hollin 2001; Ferguson et al, 1994; Mason & Gibbs, 1993; Mizuno, 1999; van Ijzendoorn & Zwart-Woudstra, 1995; Ferguson et al, 2001). In some others where individual means of males and females are reported, the Standard Deviations are not presented (e.g. Gibbs et al, 1992; Basinger et al, 1995; Comunian & Gielen, 1995). From the information that is available the findings are inconclusive. Some studies report near identical Standard Deviations for both groups (e.g. Humphries et al, 2000; Ferguson & Cairns, 1996; Barriga et al, 2001). Garmon et al (1996) present means and standard deviations of males and females from a variety of age groups, their findings regarding standard deviations are mixed. Some studies do present findings indicating that males have higher variability than females (e.g. Comunian & Gielen, 2000; Krettenauer & Becker, 2001, Gregg et al, 1994; Palmer & Hollin, 1998); however, the magnitude of the difference between the males and the females in these studies is not quite as great as the difference observed in this one.

3.33 Intra-Individual Variation in Stage Use

Comparisons of the differences in ranges of moral reasoning scores revealed that all participants were demonstrating a certain amount of stage mix in their questionnaire responses. No participant was completely consistent in stage use across all items in the questionnaire; furthermore very few participants demonstrated ranges of scores corresponding to use of adjacent stages only.

Bearing in mind that to the author's knowledge this is the first time that a systematic attempt has been made to compare patterns of intra-individual variation with this sample type using SRM-SF data, it is not clear how one should interpret the above finding in theoretical terms. As reported earlier, much of the published work using alternatives to the SRM-SF has focused on between-group differences in moral reasoning rather than on intra-individual variations in patterns of moral reasoning. The tendency has been to assume the structural consistency compatible with the Kohlbergian model rather than to examine it. Furthermore, as Krebs et al (1991) point out, the scoring systems used for Kohlbergian moral reasoning measures (Colby & Kohlberg 1987) may mask underlying heterogeneity in stage use (Krebs et al, 1991:1014-1015). The SRM-SF scoring system may be similarly problematic. Whilst the item-by-item scoring approach adopted in the current study maximised the potential for introducing intra-individual heterogeneity in one respect; the treatment of "ambiguous

response units" (Gibbs et al 1992:50) to individual items was carried out with rigorous adherence to the rules set out in the SRM-SF manual. The applications of the rules potentially reduce heterogeneity as follows: Firstly, if a moral justification matches equally well with protocols disparate by two developmental levels (e.g. Transition Stage 2-3 and Stage 3) then the justification is scored at the higher of the two levels; secondly, if a justification matches protocols disparate by three developmental levels (e.g. Stage 2, Transition Stage 2-3 and Stage 3) then the justification is to be scored at the intermediate level (i.e. Transition Stage 2-3); thirdly, if a justification matches protocols disparate by more than three developmental levels (e.g. Stage 2 and Transition 3-4) then the item is deemed unscorable and excluded from analysis. It is entirely plausible, then, that the actual level of intra-individual variation in moral stage use in the current study was higher than the results of the analyses would suggest.

In many of the studies where intra-individual structural (in)consistency in moral reasoning has been explored, the focus has been on situational factors considered to be influential on moral reasoning. In some cases investigations of internal consistency have been carried out in experimental conditions where the nature of the moral dilemmas has been manipulated comparing levels of structural consistency between Kohlbergian dilemmas and real-life dilemmas (e.g. Haviv & Leman, 2002; Wark & Krebs, 2000; Krebs et al, 1997; Wark & Krebs, 1996; Jadack et al 1995; Krebs et al 1994; Carpendale & Krebs 1992; Krebs et al 1991; Walker & Taylor 1991a). The findings from these studies is that participants do not demonstrate structural consistency in Stage use across different types of dilemmas (but see e.g. Walker et al, 1987; Walker, 1989; Walker & Moran, 1991; for studies indicating intra-individual consistency in moral stage use across different dilemma types).

It is possible that the overall pattern of intra-individual variability in stage use across the SRM-SF partially results from differences in the nature of the moral issues under consideration in the questionnaire items. For example it will be remembered from the findings presented in Chapter 2 that in general all participants tended to have lower scores on the "theft" item than to other items in the questionnaire. Conversely, scores on the item "How important is it to keep promises to children" were higher in all participant groups than to other items in the questionnaire. However, the finding

that there were between-group differences in levels of internal consistency is not easily reconcilable merely with reference to disparities in the type of moral issue under consideration. Furthermore, the finding that male non-offenders showed significantly greater amount of intra-individual variation in range of stage use than both the offenders or the female non-offenders conflicts with what one would expect to find if moral reasoning performance differed as a function of external situational variables such as task conditions. Male and female non-offenders all performed the task under similar conditions as a written exercise in a classroom setting; whereas many of the offenders were interviewed on a one-to-one basis in a criminal justice setting.

As was pointed out in the introduction to this chapter, evidence of intra-individual variation in Stage use within an adolescent sample is perhaps hardly surprising if one invokes the concept of disequilibrium (see e.g. Snyder & Feldman, 1984; Thoma & Rest, 1999; Walker et al, 2001; Walker, 1991b) as helping to account for developmental progression from one moral stage and the next. Taken on its own, the finding that there was greater evidence of intra-individual variability in terms of range of stage use amongst the male non-offenders than in the other two groups could be accommodated with reference to potential differential maturational patterns between the three groups. One could relate the differences between male and female non-offenders to known gender differences in early adolescence, and could argue that some of the younger males were exhibiting a growth-spurt previously experienced by at least some of the females. One could relate the differences between the offenders and male non-offenders to the concepts of "Structural cognitive deficit" and / or "developmental lag", perhaps by arguing that the comparatively flatter patterns of variation amongst the offenders was indicative either of global fixation at lower levels of moral reasoning or of maturational delay. However, had the between-group differences in intra-individual variation been the result of differential maturational patterns, one would have expected to have found that a greater proportion of male non-offenders were classified as exhibiting positive skew than either offenders or female non-offenders. Conversely, in the current study, no significant between-groups differences were found in the overall proportions of positive skew, no skew and negative skew.

The findings obtained from the comparisons of overall patterns of positive skew and negative skew pose problems when seeking to reconcile the differential levels of intra-individual variation in developmental terms. According to the consolidation / transition model (Snyder & Feldman, 1984; Walker & Taylor, 1991; Walker et al, 2001) the pattern of stage use that is considered to be indicative of disequilibrium is characterised both by a high level of stage mix and positive skew. In this sample, however, the overall distribution of stage use seemed to be characterised by relatively high stage mix coupled with negative skew. Very few participants restricted their stage use to adjacent stages only, and a far greater proportion of participants (approximately 61%) were exhibiting negative skew in terms of their pattern of stage use around the mode than positive skew (approximately 30%). Furthermore, when investigating between-group differences in intra-individual variability amongst those categorised as exhibiting positive and negative skew a pattern emerged that was contra to theory. Slightly higher intra-individual variability was found amongst those demonstrating negative skew rather than those demonstrating positive skew.

It was noted when presenting findings relating to the modal stage use by participant groups that when bimodality was found, the higher of the two modes was selected for use in analysis. The theoretical justification for doing this is that this a) fits in with the general acknowledgement by Kohlberg and others that people do not always perform at their optimum level of moral reasoning (see e.g. Colby & Kohlberg 1987:7-8) and b) it is in accordance with one of the SRM-SF ambiguous response unit rules that the higher of two adjacent moral levels is the preferred stage (Gibbs et al, 1992:50). However, depending on the level of bimodality in the sample, adopting this strategy clearly has the potential to bias the sample towards exhibiting negative skew – the characteristic bias that is indicative of "consolidation" (associated with low stage mix) rather than "transition" (which is associated with high stage mix). Whilst no attempt was made in the current study to investigate the frequency of bimodality within the sample, it does seem plausible that where there is relative high intra-individuality in stage use (as there seemed to be in this sample), there is also a greater likelihood of bimodality. It is not clear from some of the previous studies that have investigated patterns of consolidation and transition how this potential confound has been addressed.

Thoma & Rest (1999), using the DIT investigated patterns of stage distribution amongst participants that they had defined as being consolidated reasoners. They found little evidence of bimodality amongst this subset of their sample (Thoma & Rest, 1999: 327-328); however, they did not investigate the proportions of positive and negative skew within the sample as a whole (Thoma & Rest, 1999: 332). Walker et al, (2001) using the MJI derived their modal scores from percentage usage at each stage (Walker et al, 2001:190). They reported approximately equal proportions of positive and negative skew (Walker et al, 2001:191), but make no mention of whether there was any evidence of bimodality in their sample. Furthermore, the sample size in the Walker et al (2001) study was considerably smaller (64 participants) than that of the current study (413 participants).

Where an individual's mode falls in terms of its position on the SRM-SF scale has in any case the potential to exert bias on whether the individual is categorised as exhibiting positive skew or negative skew. It goes without saying that using the parameters of the SRM-SF, an individual who is modally "at" Stage 4 (the highest level of the scale), and is demonstrating some variation in Stage use can only be characterised as demonstrating negative skew. Equally, an individual who is modally "at" Stage 1 (the lowest level of the scale) and is demonstrating some variation in stage use can only be characterised as exhibiting positive bias in their pattern of stage response. An individual whose modal score is above the mid-point of a parametric scale with finite boundaries is more likely, *a priori* to demonstrate negative skew – there is, of course, a greater probability that a random non-modal "hit" will fall below the mode rather than above the mode. It was interesting to note from the current study when investigating the relationship between modal stage and frequencies of judgements above or below the mode that the directions of skew followed this apparent *a priori* pattern. A significant positive relationship was observed between modal stage and frequencies of judgements made below the mode, conversely a negative relationship was found between modal stage and judgements made above the mode. In the current study, then, <u>overall</u> findings relating to positive and negative skew may be a statistical artifact resulting from characteristics of the scale, rather than an indication of differential developmental states.

When comparing patterns of negative and positive skew one is presented with a confound that may not be straightforward to resolve. The direction of skew relating to the relevant modal stages is what would be predicted merely with reference to the parametric characteristics of the scale; however, clearly, it is also what one would predict from developmental theory. One would expect amongst an adolescent sample, if positive skew is predictive of developmental change, that there would be greater evidence of positive skew amongst those reasoning at the immature Stages than those reasoning at the so-called mature stages. Bearing in mind that Gibbs et al (1992) consider those modally "at" Stage 3, Transition Stage 3-4 and Stage 4 to have already attained a level of moral maturity one would therefore predict that those at Stage 3 and Transition Stage 3-4 would in any case be more likely to be consolidating existing moral reasoning than to be in a transitional shift. Clearly, one would also predict according to developmental theory and as a result of the findings presented in Chapter 2 that there would be a significant relationship between age and modal stage. Therefore, one would predict to find positive relationships between age, modal stage and number of judgements falling below the mode. Taking the sample as a whole, a significant relationship was found between age and modal stage; however no relationships were found between age and range of stage use, neither were there any apparent relationships between age and patterns of directional skew as measured by frequencies of non-modal judgements falling above or below the mode.

The current study was cross-sectional - one would need to carry out a longitudinal study using the SRM-SF exploring developmental trends in terms of patterns of intra-individual variation in order to clarify some of the issues raised above. Furthermore, as far as the author is aware, none of the currently available empirical research using the SRM-SF has explored patterns of intra-individual variation. Therefore the findings obtained in the current study are not directly comparable with previous work in this field. However, on the basis of the evidence from this sample, it does seem that there may be problems reconciling intra-individual variation in SRM-SF scores with reference to the developmental consolidation / transition model. Further work is needed in this area.

3.34 Issues relating to Global Moral Stage Status Ratings:

In the introduction to this chapter some concerns were expressed about the appropriateness of assigning Global Stage Status ratings to SRM-SF data. It was noted that the Global Stage Status point boundaries had originated from an earlier moral reasoning measure where the summary index represented an average over eight sociomoral norms; it was not clear how these boundaries could be meaningfully applied to a questionnaire containing eleven items. It also appeared that issues relating to variations in numbers of scorable responses supplied and intra-individual disparity in patterns of stage use had not been adequately addressed when considering the validity of using Global Stage Status ratings to summarise SRM-SF data. Finally, the broader point boundaries encompassing the Stage 2 and Stage 3 categories relative to the other categories in the Global Moral Stage Status Scale were considered potentially problematic, particularly in terms of potentially masking heterogeneity in stage use.

From the evidence of the findings in the current study it became apparent that the concept of Global Stage Status as it stands may be of limited use as a frame of reference to characterise SRM-SF data. On the face of it, the initial observations from between group comparisons of Global Stage Status ratings gave the impression that the offender group was characterised by a predominance of Stage 2 reasoning, whereas the non-offending group was characterised by a predominance of Stage 3 reasoning. Offender overall mean scores corresponded to a Global Stage Status rating of Stage 2, mean scores of both the male non-offenders and female non-offenders corresponded to Global Stage ratings of Transition 3(2). When examining proportions of participants "at" the original levels of the GMSS scale it was found that the modal GMSS level for offenders was Stage 2, whereas for the male non-offenders the modal GMSS level was Stage 3 and for the female non-offenders modal levels were tied at Stage 3 and Transition 3(2). However, on further investigation it was found that the picture regarding the relationship between Global Stage Status ratings and actual proportions of Stage use was far less straightforward.

Trying to apply the GMSS levels with reference to proportions of whole number of item responses was problematic. As pointed out earlier, part of the problem seemed to stem from the fact that the

SRM-SF is an eleven item questionnaire. Whilst it is not immediately obvious why Gibbs et al (1992) fixed on eleven as the number of items that were to be included in the SRM-SF (some further issues relating to the number of questions in the SRM-SF are raised in the next chapter), it is clear that having done so the GMSS point boundaries should be meaningfully applicable to the number of items in the questionnaire. This would seem to be essential when considering using the SRM-SF as an assessment tool / outcome measure in a longitudinal study and / or moral reasoning intervention, especially where global shifts in overall SRM-SF means may be slight.

One could, of course, tweak the point boundaries of the GMSS scale so that they are more meaningfully applicable to an eleven item questionnaire; however, this still leaves one with the problem of how to tackle variations in the numbers of "scorable" responses supplied to the SRM-SF. It will be recalled that all questionnaires eliciting at least seven scorable responses are considered valid for analysis purposes. In the current study, the full range of seven to eleven scorable responses was found amongst participants, with less than 12% of the total sample supplying scorable responses to all eleven items. The modal number of scorable responses supplied to questionnaire items was nine out of eleven; however, this response pattern only characterised 30% of the overall sample. Thus, using "nine scorable responses" as a benchmark would seem to be inappropriate. Furthermore, it is not clear from the previous research whether the level of variation in the proportion of scorable responses observed in the current study is a typical response pattern. As far as the author is aware, there is no currently available published information in the SRM-SF literature that specifically addresses this issue. When attrition rates are discussed in the research, they are discussed with reference to proportions of valid questionnaires supplied by participants.

There is in an additional potential problem in applying GMSS levels to SRM-SF data that is associated with the scoring procedure of the questionnaire. This problem may have been overlooked in the original research on construction and validation of the measure. As reported earlier, the use of global stage ratings to characterise patterns of moral reasoning dates back to former work using moral reasoning measures such as Kohlberg's MJI. If one assumes a relative degree in structural consistency in moral reasoning, then a GMSS scale that relates to percentage stage mix is

theoretically coherent when one considers the scoring procedure in instruments such as the MJI. Typically, in the MJI, overall scores are calculated with reference to varying percentages of major and minor stage "mix" demonstrated in response to a series of probe questions for each of a number of moral dilemmas. Typically, each moral dilemma in the MJI will elicit more than one type of moral justification and –issues regarding the treatment of anomalous justifications notwithstanding-the weighted score for each dilemma reflects this. However, in the SRM-SF, each item is coded as being at a discrete moral-reasoning level. Scores for each item are allocated according to the highest stage response unit, regardless of the number of separate moral justifications elicited for each issue under consideration. Quite aside from any issues associated with the potential loss of information involved in adopting this scoring procedure, applying a scale previously used for the essentially analogue data of the MJI to SRM-SF data may be questionable. Whilst Gibbs et al. (1992) found "acceptable concurrent validity between the SRM-SF and the MJI" (Gibbs et al. 1992:39), it is not clear from the original work validating the SRM-SF (e.g. Basinger et al, 1995) whether typical response sets associated with GMSS levels were explored. This would seem to be an issue requiring further attention in the research.

From the evidence of the current study, the wider point boundaries originally allocated to Stage 2 and Stage 3 in the GMSS scale did seem to have the effect of masking between-group heterogeneity in patterns of Stage use. The creation of sub-categories "2a" and "2b" for GMSS level Stage 2 and "3a" and "3b" for GMSS level Stage 3 revealed some between-group differences that were not identifiable when using the original scale. This was perhaps most striking when comparing gender differences amongst non-offenders whose mean scores corresponded to GMSS level Stage 3. Whilst the proportions of male and female non-offenders assigned to Stage 3 were practically identical, there were clear between-group differences in the proportions of those assigned to "Stage 3a" and those assigned to "Stage 3b". Higher proportions of males than females were assigned to the category "3b". This category had as its point boundaries mean scores ranging from 301-325; thus, those in this category had necessarily demonstrated some evidence of reasoning at above Stage 3, i.e. had given at least one moral reasoning justification corresponding to either Transition Stage 3-4 or Stage 4. Whilst the actual patterns of distribution observed in the current study may be idiosyncratic, these findings do

have implications for researchers using the SRM-SF in terms of the "justice / care" debate. As reported earlier, there is little evidence from the available SRM-SF literature to suggest that overall, males are using more Stage 4 reasoning than females. However, it may be the case that use of the broad category boundaries for GMSS level "Stage 3" hampers the identification of underlying trends of this nature.

As has previously been noted, the validity of summary mean scores and / or GMSS ratings to characterise moral reasoning patterns lies in theoretical premise that individuals are relatively consistent in their patterns of stage use. However, on the basis of the evidence from the current study it would appear that a substantial minority of the participants, especially male non-offenders, were not demonstrating the required structural internal consistency for either mean scores or GMSS ratings to accurately profile their patterns of stage use. Much of the work using moral reasoning measures presents findings with reference to these summary indices; however, in many studies little attention has been paid to sources of underlying variability in the sample. It seems clear that this lack of attention to variability may call into question some of the previously published research findings. Ideally, one would want to explore this issue further – for example through the use of meta-analytic techniques. However, as pointed out earlier, even within the SRM-SF literature, Standard Deviations around the means are not always reported. Furthermore, as Thoma & Rest (1999) point out, it is only relatively recently that researchers have turned their attention towards sources of intra-individual variation in patterns of stage use. Therefore, it may be difficult to get an accurate gauge of how this issue may have impacted on prior research. It seems imperative, then, that in future empirical studies using moral reasoning measures researchers take careful account of potential sources of variation before interpreting their findings with reference to mean scores and / or GMSS ratings.

3.35 Between-Group Differences in Proportions of "Scorable Responses" Supplied by Participants

Between group differences in proportions of scorable responses provided by participants were not statistically significant. Quite aside from considerations relating to the validity of Global Stage Status Ratings this was in any case a useful finding to establish. It will be remembered that the majority of offenders were interviewed on a one-to-one basis and that all the non-offenders performed the SRM-SF as a written task. On the face of the above findings it would appear that differences in the way the

task was carried out had no major impact on the proportion of scorable responses that was supplied by participants. However, the results of the Pearson correlations comparing the relationships between the numbers of scorable responses and overall SRM-SF scores gave an indication that the amount of verbal information supplied by males in their questionnaire responses – whether oral or written - may have been more of a factor accounting for overall performance on the measure than it was for females. Significant positive relationships between proportions of scorable responses supplied and overall SRM-SF scores were found for both the offenders and the male non-offenders. This relationship was not found amongst the female non-offenders. It may have been the case that amongst males who were supplying lower numbers of scorable items there was a greater propensity to supply short word responses corresponding to Stage 1 reasoning than there was amongst females. It would not be feasible within the remits of the current study to check if that actually was the case. Even if it was feasible to do this it is not clear how one would interpret the findings, especially as no significant relationships were found between age and proportions of scorable responses supplied by any of the participant groups. In a study carried out by Galotti et al (1991) where males and females were asked to carry out an open ended essay task giving examples of personally relevant moral issues it was found that females tended to respond more thoroughly to the task and provide more examples of moral themes than males. However, Galotti, et al (1991) were unclear as to whether this could be explained by greater cognitive complexity amongst females or whether females were demonstrating greater compliance with the task. Furthermore, it follows from Kohlbergian theory that cognitive complexity and relative conscientiousness towards completion of a task are not independent of one another when considering moral development.

3.36 Between group differences in patterns of Stage use: comparisons of offenders and non-offenders

Broadly speaking, results of the analyses comparing overall patterns of Stage use between offenders and non-offenders revealed predictable differences that were comparable to those findings presented in Chapter 2. Results of the binary Chi-Square analyses comparing offenders and non-offenders found that a significantly greater proportion of non-offenders made use of mature stages than non-offenders; the results of the chi-square analyses comparing greater or lesser use of each stage revealed that non-offenders were making more frequent use of the mature stages than offenders. However, it

was interesting to note that whilst fewer offenders than non- offenders gave mature justifications, and made less consistent use of mature stages than the non-offenders; the majority of offenders (74.1%) managed to give a Stage 3 response to at least one item in the questionnaire. Thus, whilst the offenders were apparently performing less well than offenders in terms of producing mature moral justifications, the support for the claim that offenders demonstrate an overall cognitive deficit in terms of moral development was, at best, weak. It was clear on the basis of the evidence from this sample that the majority of offenders were able to access moral maturity by providing at least one Stage 3 response to an item in the SRM-SF.

3.36i Between-group differences: comparisons of male and female non-offenders:

The results of the analyses presented in this chapter identified some clear gender differences between male and female non-offenders. A greater proportion of male non-offenders gave Stage 4 justifications than female non-offenders; a greater proportion of female non-offenders gave Stage 3 justifications than male non-offenders. Comparisons of patterns of positive and negative skew also revealed gender differences amongst those who were modally at Stage 3, with greater proportions of males than females demonstrating positive skew.

In this sample at least, there was some evidence to suggest that there were gender differences in the patterns of mature stage use that correspond to the Gilligan (1982) hypothesis. However, it should be noted that whilst "Stage 4" use was primarily restricted to the male non-offending sample, this style of reasoning was no means typical of this group. It can be seen that only 28% male non-offenders used this stage, this compares to 7% female non-offenders and 3% offenders. In contrast, whilst there was greater evidence of Stage 3 reasoning among female non-offenders than male non-offenders, high proportions of both males (83%) and females (92%) used this stage.

Comparisons of immature Stage use between male and female non-offenders also revealed some interesting gender differences, most notably in patterns of Stage 1 use. Significantly higher proportions of males used Stage 1 than females. From the findings obtained in the current study, then,

it would seem that males were making greater use of Stages at the extreme ends of the SRM-SF scale than females. Clearly, this would in part account for the failure to identify differences in the overall means of male and female non-offenders.

3.37 Item related Findings: Comparisons of Offenders and Non-Offenders

By and large, the results of the analyses comparing between-group differences in patterns of stage use amongst offenders and non-offenders across SRM-SF items corroborated the findings relating to overall patterns of stage use. All SRM-SF items managed to discriminate between offenders and non-offenders; in general, where differences were found, there was greater evidence of immature stage use amongst offenders and greater evidence of mature reasoning amongst the non-offenders. There were, however, some apparent differences in the levels of effectiveness of the various SRM-SF items to discriminate between groups. For example, whilst items relating to offending behaviour discriminated between groups for the majority of the individual stage comparisons, the items "Saving the life of a stranger" and "Living even if you don't want to" (Items that make up the "life" value) only discriminated between offenders and non-offenders for two out of a possible seven stage comparisons.

3.37i Item Related Findings: Comparisons of Male and Female Non-Offenders

As with the findings obtained from comparisons of overall patterns of stage use, some gender differences emerged when investigating between-group differences in patterns of Stage use across SRM-SF items. Between-group differences were found for the majority of SRM-SF items, these findings contrast with the findings obtained when comparing between-group differences in item means using ANCOVAs where gender differences were found for two items only. When comparing between-group differences in patterns of mature stage use, differences were most noticeable for patterns of Stage 4 use, with males demonstrating significantly higher use of this stage on five items. However, comparisons of Stage 3 use were less conclusive in demonstrating a contrasting bias towards females – differences favouring females in this respect were found on three items only. Comparisons of Transition Stage 3-4 use revealed differences for two items only, in both cases significantly higher proportions of males used this stage than females. Comparisons of between-

group differences of immature stages also revealed some gender differences: males demonstrated greater use of Stage 1 than females on two of the items; females demonstrated greater use of Transition Stage 2-3 on four of the items.

3.38 Problems interpreting the findings relating to Transitional Stages.

Whilst use of categorical techniques enabled one to uncover some between-group differences in patterns of stage use that were not readily identifiable with reference to summary means; there were problems when it came to interpreting the findings relating to the Transitional Stages (i.e. Stages 1-2, 2-3 & 3-4). It will be recalled from the overview to the SRM-SF that each major moral Stage may be further sub-divided into content categories known as "Aspects" (for example, Stage 3 has six different aspects: "Relationships", "Empathic Role-Taking", "Normative Expectations", "Prosocial Intentions", "Generalized Caring", and "Interpersonal Approval"). Broadly speaking, the Transitional Stages in the SRM-SF do not in themselves represent discrete types of moral reasoning, but are combinations of Aspects from adjacent major stages. (Transition Stage 3-4 is, of course somewhat different in this respect as whilst it can be seen as a mid-way point between Stage 3 and Stage 4, there is a distinct style of reasoning - "relativism of personal values"- associated with this stage).

As Gibbs, Basinger & Fuller (1992) put it:

"The transitional phases generally do not lend themselves to distinct discussion because their content categories are blends of aspects of the adjacent stage levels" Gibbs, Basinger & Fuller, 1992:28

So, for example, each Aspect in Transition Stage 2-3 combines Aspects from Stage 2 ("Exchanging and Instrumental") and Stage 3 ("Mutual and Prosocial"). In global terms Transition Stage 2-3 is seen as a move away from "reciprocity as fact" and a shift towards "more hypothetical and intrinsic aspects of sociomoral behaviour". However, it is not clear from Gibbs, Basinger & Fuller (1992) whether one should interpret Stage 2-3 as being tending more towards Stage 3, more towards Stage 2 or as an equal mix of both Stages. Intuitively, if one made reference to theory and GMSS levels, one would categorise Transition Stage 2-3 as more representative of Stage 3 than Stage 2: the score allocated to

Transition Stage 2-3 is 250, this falls at the lower end of Transition 3(2) in GMSS terms, which is deemed to represent "majority Stage 3 usage and minority Stage 2 usage". However, given the problems associated with GMSS levels described above, this type of characterisation would be purely speculative.

It was stressed in the introduction to this thesis that the findings relating to gender differences presented in this document were restricted to the investigation of structural differences in moral judgement (or differences in global stage usage) and that the issue of between-group differences in moral content or orientation within global stages was not one that would be addressed in the current study. Given the problems associated with interpreting the findings relating to the Transitional Stages with reference to structure alone, clearly one would wish to turn to content analysis of SRM-SF protocols as a means of resolving the issue. However, analysis of Transitional Stage protocols with reference to the Aspects or content categories of the SRM-SF is in itself potentially problematic.

As noted earlier, the "Aspects" in each Transitional Stage are combinations of Aspects from adjacent major stages. For example, for the "Property and Law" sociomoral value Transition Stage 2/3: "Aspect 6" combines the "Needs" aspect of Stage 2 with the "Empathic Role-Taking" aspect of Stage 3. This combination could be theoretically coherent in terms of representing a specific style of reasoning when one considers the Bear and Rhys (1994) argument that Stage 2 "fuses" both a "hedonistic" style of moral reasoning and a more empathic "needs" based style of moral reasoning. One could argue that this particular combination would represent a care focus rather than a rights or justice focus. However, some of the other Transitional "Aspects" are not as straightforward to interpret in theoretical terms. For example, Transition 2/3: "Aspect 8" combines "Advantages" (Stage 2, Aspect 6) with "Prosocial Intentions" (Stage 3, Aspect 4) and "Empathic Role Taking" (Stage 3, Aspect 2). On the face of it, this combination incorporates both care and justice orientations.

A further difficulty in analysing SRM-SF data by content category emerges when one considers that a single SRM-SF item may elicit more than one moral justification. Whilst there are rules in the manual concerning the scoring of justifications that correspond to different moral stage levels, no

mention is made of how one should deal with protocols that correspond to different aspects within stages. It should also be noted that the way the Aspects are combined and the number of content categories in the Transitional levels vary according to the sociomoral values under consideration. The examples given above come from the "Property and Law" value which has ten content categories or sets of combined Aspects; however, if one turns to the "contract and truth" sociomoral value, one finds that there are only seven content categories or sets of combined Aspects. So, for the "Property and Law" value one finds that Transition Stage 2-3, Aspect 4 is a combination of "Freedoms" (Aspect 3 from Stage 2) and "Normative Expectations" (Aspect 3 from Stage 3). However, no such combination exists for the "contract and truth" value. Furthermore, there is in any case, no content category for Transition 2-3 "contract and truth" that incorporates "Normative Expectations" – this despite the fact that the "Normative Expectations" Aspect is represented in the Stage 3 protocols for this sociomoral value.

The difficulties associated with making sense of the Transitional Stages are particularly problematic when considering the Transitional Stages 2-3 and 3-4 (save for the items associated with offending behaviour, little use was made of Transition Stage 1-2 by any of the participants). The majority of all participants – including the offenders – were demonstrating evidence of Stage 2 and Stage 3 mix to some degree, and the modal Stage of both the offenders and the female non-offenders was Transition Stage 2-3. In this sample, the comparative immaturity of the offenders compared to non-offenders appeared to be a matter of levels of performance rather than due to an innate cognitive deficit. Clearly, one would wish to explore patterns in styles of reasoning further, for example to try and identify when and where offenders are most different from their non-offending peers. However, the confusion surrounding the interpretation of Transition Stage 2-3 in terms of both structure and content has resulted in a loss of potentially valuable information in this respect.

It follows from the above that when considering the interpretation of the findings relating to Transition Stage 3-4 there are problems in differentiating between "justice" and "care" orientations within this particular Stage. As with Transition Stage 2-3 some of the combinations of "Aspects" seem to be more theoretically coherent than others. For example, in the property and law sociomoral

value one finds that Transition 3-4: that "Aspect 4" combines "Normative Expectations" (Stage 3, Aspect 3) with "Societal Requirements" (Stage 4, Aspect 1). This combination would seem to be well-matched in terms of representing an extension of interpersonal norms towards the requirements of a complex social system (Gibbs et al, 1992: 29). However in some other content categories the Aspects that have been combined are if anything conflicting rather than complementary. For example, "Aspect 2" combines "Empathic Role-Taking (Stage 3, Aspect 2) with "Basic Rights or Values" (Stage 4, Aspect 2), and "Aspect 6" combines "Normative Expectations" (Stage 3, Aspect 3) with "Responsibility" (Stage 4 Aspect 3).

Given that Gilligan differentiates between the "rights oriented morality" of males and the "responsibility oriented morality" of females (Gilligan 1982:22); the Aspect 2 match of an "empathic" (strong care focus) aspect from Stage 3 with a "rights" (strong justice focus) aspect of Stage 4 seems rather odd. Intuitively, from the above examples, it would seem to be more appropriate to combine "Empathic Role-Taking" with "Responsibility" and "Normative Expectations" with "Basic Rights or Values".

It would seem, then, that when seeking to analyse moral reasoning using categorical interpreting findings relating to the SRM-SF "Transitional Stages" may be problematic in terms of uncovering potential sources of heterogeneity in the sample. This issue does not seem to be easily reconcilable with reference to the individual content categories of these Stages as they currently stand.

One could perhaps resolve the problem by adopting qualitative techniques to code the data, However, this would then negate the use of a standard scoring manual, and, in any case might not be practically feasible with large data sets such as the one in the current study. In the introduction to this thesis it was noted that Garmon et al (1996) found no evidence to support the claim that the SRM-SF exhibited stage bias favouring males, but did find some evidence to support the claim that females were more care focussed in terms of content of moral judgements within each moral stage. However, the findings presented by Garmon et al (1996) are restricted to the consideration of differing moral orientations of those providing Stage 2 and Stage 3 justifications, and no information is given regarding moral orientations of Transition Stage 2-3 reasoners. The study carried out by Garmon et al (1996) would appear to be the only work using SRM-SF data that has sought to investigate moral

orientation with reference to the individual Aspects within Global Stages. It is, therefore, unclear how previous researchers have tackled the apparently problematic nature of interpreting findings relating to the Transitional Stages. This issue is one that requires clarification in future studies using SRM-SF data.

3.39 Recoding SRM-SF Data to Impose an Upper Limit of "Stage 3" on the Data

The final set of analyses that were carried out involved "recoding" the SRM-SF data in order to impose arithmetical homogeneity on the different "mature" levels of moral reasoning (i.e. to treat "Stage 3", "Transition Stage 3-4" and "Stage 4" as equivalent in numerical terms). The main purpose of this set of analyses was to demonstrate a methodological point: i.e. how heterogeneity in "mature stage" use might impact on summary mean scores. No claims are made regarding the legitimacy of this approach in terms of moral reasoning theory – indeed one would not wish to argue that "Stage 3" "Transition Stage 3-4" and "Stage 4" could all be treated as representative of the same global stage of moral reasoning in theoretical terms. However, as pointed out in the overview to the measure, there seems to be somewhat of a conflict in the SRM-SF literature. On one hand it is acknowledged that Stage 3, Transition Stage 3-4 and Stage 4 are all representative of moral maturity, however, for the purposes of statistical analysis these Stages are treated as representing different points on an ascending scale.

Gibbs et al (1992) attempted to address the issue of potential gender bias in measures of moral reasoning during the initial validation of the SRM-SF. Along with other researchers their assumption has seemed to have been that were there to be stage bias favouring males in the SRM-SF this would be picked up through between-group differences in summary means. As reported earlier, there is little evidence from the SRM-SF research to indicate that this is the case, indeed, where gender differences have been found on the measure typically the trend has been for adolescent females to out-perform adolescent males.

In lay terms, then, many researchers seem to be content with the view that females are not "handicapped" in performance terms on the SRM-SF; however, this is clearly not the same thing as

giving females a "level playing field" to start off with. The exploratory recoding of the data attempted to address this issue. Comparisons of the findings obtained from the recoded data with those using the original coding scheme indicated that had the females been given a level playing field to start off with then the initial picture regarding between-group differences in overall means presented in Chapter 2 would have been somewhat different.

Without falling into the trap of trying to prove the null hypothesis, it is clear that had there been no bias favouring male non-offenders in the current study one should not have observed any differences in the findings from the overall ANCOVAs according to the coding schemes used. However, in the current study, whilst there were no between-group differences in the overall means of male and female non-offenders using the original coding scheme, once the data had been recoded a gender difference emerged that favoured females. This finding was corroborated when comparing between-group differences in the means of SRM-SF items using the recoded data. Whereas when the original coding scheme was used between group differences were found for two of the items "Telling the Truth" and "Saving the life of a friend" (with males having significantly higher means on the item "Telling the Truth" and females having significantly higher means on the item "Saving the life of a friend"), once the data were recoded between-group differences were found favouring females on three additional SRM-SF items.

On the face of it, recoding the data did little to change the picture regarding the differences between offenders and non-offender overall means. For both sets of analyses between-group comparisons of offender / male non-offender means were statistically significant at the $p<0.001$ level. However, it was interesting to note that recoding the data appeared to have some impact on the magnitude of the F value for the offender / male non-offender comparison. When using the original coding scheme the offender / male non-offender F value was 111.49 (1,293), whereas once the data were recoded the F value dropped to 86.94 (1, 293). As with the overall mean comparisons, at first sight the recoding of the data did little to alter the picture regarding the individual SRM-SF item comparisons. For the offender / male non-offender comparisons both sets of analyses identified statistically significant differences for all eleven items. However, on further investigation, here too there was some indication

to suggest that recoding the data did have an impact on the magnitude of the offender / male non-offender comparisons. When the original coding scheme was used between group item comparisons of offenders and male non-offenders found significant differences at the $p<0.001$ level for nine of the eleven items and significant differences at the $p<0.01$ level for the remaining two items; however when the data were recoded the number of significant differences at the $p<0.001$ level had dropped to seven out of eleven, with two of the remaining items achieving significance at the $p<0.01$ level and two achieving significance at the $p<0.05$ level.

3.4 CONCLUSION

The accumulated evidence from the various sets of findings presented in this Chapter corroborated the view expressed at the end of Chapter 2 that reliance on mean scores as the sole of information about the moral reasoning of adolescents is problematic. Furthermore, results of the subsequent analyses cast serious doubt on the validity of GMSS ratings to characterise the SRM-SF data set in the current study.

Bearing in mind that these two summary indices are the ones that are most commonly presented in the moral reasoning literature, these findings may have potentially serious implications – both in terms of interpreting the previous research in this field and for future work in the field. It became clear during the course of the analyses that inter-individual and intra-individual variation in moral stage use within samples are factors that need much closer attention in any future SRM-SF research.

Heterogeneity in mature stage use uncovered by the analyses presented in this chapter seemed to be of primary importance when considering the potential "distortion" of the original picture regarding between-group differences in SRM-SF mean scores. From the results of the categorical analyses carried out in the current study it was clear that this distortion was largely attributable to presence of Stage 4 reasoning within the male non-offending sample not evident amongst the other two groups (it will be recalled that amongst the offenders and female non-offenders use of Stage 4 reasoning was negligible). This disparity seems to have "inflated" the moral maturity of the male non-offenders relative to the other two groups. The effect was most noticeable when comparing gender differences amongst the non-offending sample; however, the potentially distorting impact of this type of disparity on offender / male non-offender comparisons should not be overlooked.

The final chapter that follows attempts to bring together some of the general themes that have emerged from the study as a whole. The key outcomes of the study are presented and the overall findings will be discussed with reference to their more global theoretical contexts.

CHAPTER 4: GLOBAL DISCUSSION

4.1 Overview

The current study set out to address a number of theoretical and methodological issues concerning the assessment of "moral reasoning" in adolescents using the SRM-SF. In the first section of this final chapter, the main findings to emerge from the empirical studies presented in Chapter 2 and Chapter 3 are reviewed. In the second section issues relating to the findings are discussed in their theoretical context. Some suggestions for future research are also provided. Concluding remarks are presented at the end of the chapter.

4.11 Review of Analyses carried out in Chapter 2:

The analyses presented in Chapter 2 included a partial replication of previous work using the SRM-SF to compare between-group differences amongst offending and non-offending adolescents. It was considered to be essential to present this replication before carrying out further studies using alternative analytic techniques. By and large, initial findings comparing between-group differences corroborated previous work in this area – with male non-offenders having significantly higher overall means than the offenders. Comparisons of gender differences amongst the non-offenders found no significant differences in overall SRM-SF means. Broadly speaking, comparisons of individual item means replicated the findings obtained from the analyses of summary mean scores.

A variant to previous work was that age of the participants came more closely under scrutiny than in some studies that have used similar participant groups. Participants were categorised according to three different age bands and separate analyses were repeated for age group. The key differences to emerge as a result of classifying participants according to age related to the youngest age band. In this category no significant differences were found in the means of offenders and male non-offenders; however, female non-offenders demonstrated significantly higher means than the male non-offenders. As with the participant group as a whole, results of between-group comparisons of individual item means within age bands corroborated the findings obtained from SRM-SF summary means. Comparisons of SRM-SF overall means across age-group categories were also carried out. The findings from these set of analyses indicated that offender mean scores showed little change across

age group bands, whereas non-offenders' mean scores rose by age. Amongst the non-offenders the "improvement" in SRM-SF means across age categories was more pronounced for males than females.

The variability around the SRM-SF means was explored. Perusal of the Standard Deviations around summary means revealed that the Standard Deviation was somewhat higher for the male non-offenders than for the other two participant groups. Comparisons were made between the Standard Deviations obtained in the current study and those presented in reported findings in the SRM-SF literature. It was found that whilst the magnitude of the difference between the Standard Deviation of the male non-offenders and the other two groups was higher than that quoted in some previous studies, it was by no means aberrant. However, it was noted that it was difficult to make more than speculative claims regarding the comparability of the Standard Deviations observed in the current study, as information of this nature is not always reported in the SRM-SF literature. It was posited that heterogeneity in use of mature stages of moral reasoning may have been a factor accounting for the greater variability amongst the male non-offending sample.

In addition to replicating previous SRM-SF comparative studies of offending and non-adolescents, some tests of the internal consistency of the SRM-SF were also carried out. These included inter-item correlations, related t tests on items within the so-called sociomoral values and Factor analyses on SRM-SF items. Whilst the results of all inter-item correlations were statistically significant, the results of the other analyses were less conclusive in terms of demonstrating internal consistency of the scale. The results of some of the related t tests posed questions about the validity of the sociomoral values as representing coherent content categories. The results of the factor analyses raised some issues regarding the structural unity of the SRM-SF domain.

The final set of analyses to be presented in Chapter 2 explored the nature of the decisions participants made about the "importance" of each moral issue under consideration (i.e. whether participants rated an issue as very important, important or not important). Whilst the majority of the participants rated

all SRM-SF issues as important, differences between offenders and non-offenders were significant on a number of items with higher proportions of non-offenders rating issues as unequivocally important.

4.12 Review of Analyses Presented in Chapter 3:

The bulk of the between-group analyses that were presented in Chapter 2 had been carried out with reference to summary mean scores of either the questionnaire as a whole or of individual SRM-SF items. However, in the Conclusion to Chapter 2, concern was expressed about the efficacy of overall summary means to adequately represent moral stage use within the sample. The subsequent analyses that were presented in Chapter 3 sought to address this issue by investigating patterns of Stage use through the use of alternative methods of analysis.

The first set of analyses set out to investigate the properties of the GMSS scale. The GMSS scale was of interest as GMSS ratings are commonly used in the moral reasoning literature to characterise the moral reasoning of participant groups. However, given that GMSS ratings are based on summary mean scores and the concerns that were expressed about the mean as a result of the analyses that were presented in Chapter 2, it was felt that the reliability and validity of the GMSS scale needed investigation. Furthermore, it was noted that the GMSS scale had been derived from previous moral reasoning measures but had effectively been directly transferred onto SRM-SF data; therefore the applicability of this scale to characterise SRM-SF data might be questionable. As a result of the investigations into the properties of the GMSS scale, the conclusion was drawn that GMSS ratings as they stood were of limited use to characterise SRM-SF data, especially amongst a sample where there was evidence of variability in stage use.

Between-group differences in modal Stage use were also investigated. Comparisons of between-group differences of offenders and male non-offenders were statistically significant; the differences were mainly attributable to higher proportions of non-offenders than offenders being modally "at" higher stages of moral reasoning. Comparisons of male and female non-offenders did not reveal any significant differences. However, it was noted that whilst modal stage was perhaps superior to the

mean in terms of representing actual patterns of Stage use, the reliability of the mode as a summary index was equally susceptible to variability in the sample.

A key aim of the analyses presented in Chapter 3 was to explore potential sources of variation in stage use within the sample and to investigate the potential impact that this may have had on the reliability of various summary indices. Variation in the sample was investigated from a number of different perspectives. Comparisons of both intra-individual and inter-individual variation in Stage use were carried out. Patterns of intra-individual variation were of particular interest, as this seemed to be an area that had been largely neglected in previous SRM-SF research. Overall, there was evidence of relative intra-individual variability in patterns of Stage use, with few participants demonstrating a range of stages restricted to adjacent levels only. Between-group differences in were also found – it emerged that male non-offenders as a group were less structurally consistent than either the offenders or the female non-offenders.

Having identified that there was relative intra-individual variability in patterns of Stage use, some attention was paid to identifying the potential sources of the variation. From a literature review, it appeared that aside from gender, two other factors figured prominently in the moral reasoning research into this area. One of these related to external "situational variables" potentially influential on moral reasoning performance, for example the nature of issue under consideration and/or the nature of the task condition. The other prominent factor related to internal mechanisms of developmental change - most notably with reference to so called patterns of "consolidation and transition" in stage use.

On further investigation it was found that situational factors were unlikely to be the sole cause of variability in the sample. Whilst variations in the nature of the issue under consideration undoubtedly impacted on the overall variability in the sample (for example, participants tended to perform rather better on the item relating to keeping promises to children than to other items in the questionnaire, and performed less well on the "theft" item), this would not have accounted for the between-group differences in variability that emerged. Variation appeared to be unrelated to task condition in terms

of number of scorable responses supplied by participants. The finding that male non-offenders were the group who were demonstrating highest intra-individual variation also called into question a "situational" interpretation of the findings. Male and female non-offenders had all performed the task under similar conditions – as a classroom exercise - whereas many of the offenders had been interviewed on a one-to-one basis.

Applying the "consolidation and transition" model to the overall SRM-SF data proved problematic. A particular difficulty involved trying to relate patterns of modal skew to levels of intra-variability in the sample. Overall, whilst there was evidence of intra-individual variation in stage use, the sample was negatively biased in terms of modal skew. This seemed contra to theory. It was acknowledged that the selection of the "highest mode" when bimodality was present may have been partially responsible for this finding. However, it was pointed out that when intra-individual variability increases the risk of bimodality does also; it was not clear how previous studies had tackled this issue.

It was noted that there were potential artifactual confounds associated with the "consolidation and transition" model related to inherent characteristics of moral reasoning scales; as with the question surrounding bimodality it was not apparent whether these issues had been considered in previous work. However, whilst there appeared to be problems applying the "consolidation and transition" model to the overall data set, it was felt that between-group differences in "positive" and "negative" skew around modal levels "Stage 2", "Transition Stage 2-3" and "Stage 3" would be potentially revealing and worthy of further investigation.

A further aim of the analyses presented in Chapter 3 was to explore between-group differences in patterns of stage use using categorical techniques. A recent meta-analysis into gender differences in moral reasoning (Jaffee & Hyde, 2000) had concluded that whilst overall there was limited evidence to support the claim that males and females differ in terms of justice and care focus in moral reasoning, certain factors appeared to moderate the magnitude of these differences. Differences were apparently greatest amongst adolescents, when using measures that did not make use of a dilemma and when the outcome was categorical rather than continuous (Jaffee & Hyde 2000:720).

Furthermore, Jaffee and Hyde had expressed the need for further work in this field using measures that could be scored using both continuous and categorical techniques.

Comparisons of patterns of stage use amongst offenders and non-offenders revealed findings comparable to those obtained when comparing overall SRM-SF means in Chapter 2. Non-offenders demonstrated more consistent use of "mature" stage use than offenders across SRM-SF items. However, despite the apparent moral "immaturity" of the offenders, it was felt that making the claim that offenders were demonstrating an overall structural cognitive deficit would be misplaced. Whilst non-offenders were "out-performing" the offenders on the SRM-SF, the majority of the offenders had evidenced Stage 3 reasoning to at least one item in the questionnaire.

In contrast to the findings presented in Chapter 2, clear gender differences were found in patterns of Stage use. The main source of the differences appeared to result from heterogeneity in mature stage use, with only the male non-offenders demonstrating more than negligible use of Stage 4. However, some gender differences were also found in patterns of "immature" stage use – with male non-offenders making greater use of Stage 1 reasoning. This greater use by male non-offenders of Stages at the extreme ends of the SRM-SF would, in part, account for the failure to uncover gender differences through the comparison of group means alone.

When treating the data in categorical terms a problem arose regarding the interpretation of the findings relating to the Transitional Stages. It was pointed out that the Transitional Stages were essentially "blends" of adjacent major stages, and as such could not be considered to represent distinct styles of moral reasoning. This created particular difficulties when seeking to interpret findings relating to Transition Stage 2-3 and Transition Stage 3-4 (save for on items relating to offending behaviour, use of Transition Stage 1-2 was negligible amongst participants in the current study). Furthermore, referring to the individual content categories or "Aspects" within stages did not appear to resolve this issue.

The final set of analyses attempted to demonstrate how heterogeneity in mature stage use had potentially impacted on the original summary mean scores. Concern was expressed that the greater use of Stage 4 reasoning amongst male non-offenders established through the use of categorical analyses may have inflated the overall means of this sample, and thus may have distorted the original findings regarding the relative overall "moral maturity" of the participant groups. It was noted that the general finding that females perform "at least as well as" males on the SRM-SF was not in itself evidence that the measure is gender-fair. Data were recoded to impose arithmetical homogeneity on the mature levels "Stage 3", "Transition Stage 3-4" and "Stage 4". Exploratory ANCOVAs were carried out, and findings compared with those obtained using the original coding scheme. As suspected, the findings differed according to the coding method used. The main difference to emerge related to between-group differences in the overall means of male and female non-offenders. Whereas no significant differences had been observed between the overall means of male and female non-offenders using the original coding scheme, once the data were recoded significant differences were found between these two groups, with females apparently demonstrating higher levels of moral maturity than males.

4.2 DISCUSSION OF KEY FINDINGS IN THEORETICAL CONTEXT

4.21 Issues Relating to the Scoring Procedure of the SRM-SF

It was noted in the overview to the measure that the conventional way of scoring the SRM-SF questionnaire by questionnaire rather than item by item was potentially suspect in terms of reducing intra-individual heterogeneity in Stage use. It was felt that this method potentially introduced a "rater expectancy" effect. In the current study, this issue was addressed by technique of scoring SRM-SF data item by item. However, there are other aspects of the scoring procedure which may be problematic in this respect and that are less easy to address. For example, the treatment of "ambiguous response units" and multiple justifications to a single SRM-SF item potentially results in loss of valuable information regarding heterogeneity in Stage use. It will be recalled that "ambiguous response units" that express ideas corresponding to disparate moral levels are discounted as "unscorable". When "ambiguous response units" express ideas that are equally matched at two

adjacent moral levels the justification is scored at the higher level, and when "ambiguous response units" express ideas equally matched to three disparate levels the justification is scored at the intermediate level (Gibbs et al, 1992:50). When multiple justifications are provided to a single SRM-SF item, the item is scored at the highest developmental level evidenced.

One of the main reasons that the SRM-SF was devised – aside from addressing the theoretical concerns surrounding Kohlberg's Stage 5 and Stage 6 – was to provide a more reliable and practically applicable alternative to the MJI. The main practical difficulties associated with the MJI lie with the intricacies of inferential assessment of moral protocols and the complexities of the scoring procedure (Gibbs et al, 1992). The SRM-SF has clear advantages over previous measures in terms of "user-friendliness" and inter-rater reliability; however, - despite the finding that the measure demonstrates "acceptable validity" (Gibbs et al, 1992:39) - it is possible that the emphasis on improving overall reliability has made the SRM-SF less robust than some other moral reasoning measures in terms of underlying ecological validity.

It is interesting to note that one of the criticisms that Gibbs et al (1992) make of so-called recognition measures is that the developmental level of moral justifications one recognises as being important may not correspond to the developmental level of moral justifications that are spontaneously produced. Or as Rest et al (1999) put it:

*"as production tasks probably **under**-estimate a person's development, a recognition task probably over-estimates a person's development"* (Rest et al. 1999:52).

However, the general strategy of scoring each SRM-SF item only at the highest moral level evidenced may in itself create a bias towards over-estimating a person's moral development. Furthermore, by limiting the coding procedure to the allocation of just one moral reasoning score for each item one has potentially restricted the amount of information available regarding inter-individual differences in moral reasoning performance. For example an individual who gives three "Stage 2" justifications and one Transition Stage 2-3 justification to a single item will be scored at same stage for that item – Transition Stage 2-3 - as an individual who provides four Stage 2-3 justifications to a single item. In

terms of "moral competence", both individuals are "at" the same level, however, in terms of underlying moral performance the above response profiles are very different.

It will be recalled that when comparing overall means of offenders and male non-offenders in the youngest age group category, whilst the non-offenders had higher means than the offenders, between-group differences were not statistically significant, with both groups apparently demonstrating a predominance of Stage 2 reasoning. In terms of overall moral competence, then, the two groups were apparently very similar. However it may have been the case that had one made reference to levels of performance through the consideration of multiple justifications to single items one might have uncovered differences between the two groups. Whilst no attempt was made in the current study to investigate the proportion of "multiple justifications" supplied to items, from eyeballing the data in the current study and on the basis of the author's previous experience in scoring SRM-SF questionnaires, it is apparent that British adolescents frequently supply multiple justifications to individual SRM-SF items.

In Chapter 3 the attempt to apply the "Consolidation and Transition" model to SRM-SF data was unsuccessful. This was despite the suggestion made by Walker et al (2001:196) that the "Consolidation / Transition" model usefully be applied to the measure. However, Walker et al (2001) did note that their approach is more suited to deal with all moral reasoning data produced in response to issues, rather than summary item scores:

"the Bayesian approach introduced here is designed to deal directly with the distribution of stage-typed raw scores, rather than summary scores or the SRMS, but otherwise these techniques could be readily be brought to bear on the raw data provided by either the SRM or the SRM-SF" (Walker et al 2001:196)

By the inclusion of all moral justifications provided to individual items - rather than restricting the data set to one summary score per item - one would perhaps be provided with raw data more sensitive to between-group differences and that are more amenable to analysis with reference the consolidation / transition model. However, this then leaves the question of how one would modify the SRM-SF

scoring procedure in such a way that sufficient information is provided, but that the measure retains its advantage over the MJI in terms of "user-friendliness" and "reliability". It should be noted that one of the reasons that Walker et al (2001) put forward for applying the consolidation and transition model to the SRM-SF is because it is more easily administered and scored than the MJI (Walker, et al, 2001:196).

4.22 Use of summary indices to characterise moral reasoning of participants

Use of any summary index poses problems in terms of effecting an uneasy compromise between providing a manageable way of rationalising data and losing information about individual differences within the sample. In the current study problems were identified with all summary indices commonly used to represent SRM-SF data (the summary mean, GMSS rating scale and the mode) in terms of potentially masking heterogeneity in the sample. Comparisons of overall means and modal stages failed to identify gender differences between the non-offenders that were later identified when comparing frequencies of stage use. It was found that inter-individual and intra-individual variability impacted on the utility of GMSS levels to characterise patterns of Stage use.

There is an increasing awareness in the forensic literature that when considering the impact of clinical interventions with young people there is a need to pay greater attention to individual differences within sample populations. In the current study, in line with previous work, between group differences were found in overall SRM-SF means of offenders and non-offenders in the predicted direction. However, whilst it is clearly of use to be able to establish that the measure is able to discriminate between the two groups, this finding may be of limited clinical applicability. Friendship, Falshaw and Beech (2003) and Friendship et al (2003,b) comment that in the past researchers evaluating the impact of offending behaviour programmes have over-emphasised the use of group means, statistical significance and effect size as outcome measures (Friendship, Falshaw and Beech, 2003:120) (Friendship et al 2003, b). They point out that whilst group means may give one an indication of the collective outcome of an intervention they "mask" individual outcomes and are of limited value in terms of identifying individual clinical need at the pre-test stage.

Bearing in mind that moral reasoning forms a component of the most widely run cognitive behavioural interventions in UK prisons (Blud et al, 2003; Wilson, et al 2003), one would wish to use a measure such as the SRM-SF as an assessment tool at pre-test and at post-test. However, researchers considering using the SRM-SF as an outcome measure may wish to take into consideration the potential shortcomings of summary indices to characterise patterns of moral reasoning within participant groups.

4.23 Attrition Rate Resulting from Unscorable Items

It is worth remembering that potentially valuable information regarding heterogeneity in Stage use had been already been discarded at a very early phase of the current study through the exclusion from analysis of those questionnaires resulting in fewer than 7 scorable responses. This strategy was adopted because it followed the guidelines in the SRM-SF manual:

"Questionnaires yielding fewer than 7 scorable responses do not result in reliable protocol scores, and should be discarded from analysis" (Gibbs et al 1992:55). Gibbs et al (1992) do not qualify this statement, so it is unclear on what basis this decision was made. Intuitively, one would imagine that this decision is based on the proportion of scorable information provided to the questionnaire as a whole: If an individual provides 7/11 scorable responses they have "given answers" to approximately 64% of the questionnaire items, if an individual gives only 6/11 scorable responses this percentage drops to approximately 55%. However, due to the structure of the SRM-SF, a decision made on this basis may not be as straightforward as it initially appears.

The 11 SRM-SF items are grouped under 5 content categories known as sociomoral values: "Contract and Truth" (items 1-4), "Affiliation" (items 5 & 6), "Life" (items 7 & 8), "Property and Law" (items 9 & 10) and "Legal Justice" (item 11). Leaving aside any issues relating to the internal consistency of the sociomoral values for the time being, it is clear that questionnaires deemed "valid" for analysis may vary considerably in terms of the content that has been covered. The only way to ensure that participants have provided information relating to all the moral values is if they have provided scorable responses to every SRM-SF item (this is due to the "Legal Justice" value being represented by one item only). On the other hand, clearly, it is possible to cover all the sociomoral values with

only 5/11 responses. It seems questionable to make the assumption that 7/11 scorable responses covering only 3/5 content categories is more "reliable" than 5 or 6 responses that cover all the sociomoral values.

Information about heterogeneity in the proportion of content covered in the SRM-SF is clearly of relevance to the researcher. Firstly, if there are disparities in the amount of scorable information provided to the sociomoral values, this may raise issues regarding the face validity of some of these values. For example, in the current study, overall, somewhat fewer participants provided scorable responses to the "Life" value (n=365) than to the "Affiliation" value (n=406). Secondly, variation in the proportion of SRM-SF content covered may impact on the overall means. In the current study, even though all inter-value correlations were statistically significant, participants in all three groups tended to perform less well on the "Property and Law" value (means of offenders = 192.59, means of male non-offenders = 245.86, means of female non-offenders 244.37) than on the "Contract and Truth" value (means of offenders = 228.40, means of male non-offenders = 277.57, means of female non-offenders 271.88).

It follows that the observations made regarding the sociomoral values also apply to the individual SRM-SF items. Furthermore, on the basis of the evidence from the current study, whilst all inter-item correlations were statistically significant, the results of the related t tests carried out on items within the "values" had raised some questions regarding the internal consistency of some of the sociomoral values. For example, within the "Contract and Truth" value overall mean scores were significantly lower for the item "Telling the Truth" than they were for the item "Keeping a promise to a friend".

As with the sociomoral values, between-group differences of offender and non-offender item means were more pronounced for some items than for others - for example. However, as was reported in the overview to the measure, few studies have analysed between-group differences in the SRM-SF on an item by item basis. To the author's knowledge there is only one (Palmer & Hollin, 1998) that has presented findings related to individual SRM-SF items with a broadly comparable UK based sample,

and in the Palmer & Hollin (1998) study the participants were somewhat older than the ones in the current study.

The overall attrition rate was somewhat higher in the current study than that quoted in some of the previous work. The attrition rate can in part be attributed to task conditions – not all offenders were willing or able to supply moral justifications to all items; non-offenders completed the task as a classroom exercise, and it was not possible to monitor individual response sets or to use "probe" questions when moral justifications were not forthcoming to particular items. However it was noted that when scoring the SRM-SF that four items in particular seemed to elicit a higher proportion of "unscorable" responses than others. These were: "Keeping a promise to a stranger", "Saving the life of a stranger", "Living even if you don't want to" and "Judges sending lawbreakers to jail". The strategy of scoring each questionnaire item-by-item probably accounts for the higher proportion of unscorable responses supplied to the "Stranger" items – these items "follow on" from questions that ask about "keeping a promise to a friend" and "saving the life of a friend", and a number of respondents provided answers such as "Same as before" to these items. (Whilst participants are instructed not to answer questions in this way, there are guidelines in the SRM-SF manual that allow one to assign the "same score" to this type of response as was assigned to the previous response. Clearly in the current study, having adopted the strategy of scoring responses item-by-item it would have been inappropriate to follow this procedure). However, the apparent difficulty in supplying scorable responses to the other two items is less easy to explain with reference to task conditions.

As was described in the "method" section of the current study, a high proportion of unscorable responses to the item "How important is it to live even if you don't want to?" arose as a result of participants giving answers like "it's up to them", "it's their choice" or "it's their life". It was pointed out that there was an apparent discrepancy in the SRM-SF manual relating to this type of response, in that it is given as an example of a scorable Stage 2, Aspect 3 "Freedoms" type response (Gibbs et al, 1992:100) but cited as being unscorable when presenting Transition Stage 3-4 protocols to this item (Gibbs et al, 1992:106). On the face of it, it would seem that without further qualification this type of

response is equally well matched to Stage 2 and to Transition Stage 3-4 "Relative Personal Values". For the sake of rigour, then, in the current study, this type of response was deemed to be unscorable.

The SRM-SF is more user-friendly in many respects than the MJI, however, it may be at somewhat at a disadvantage regarding the quantity of the moral reasoning data that it elicits to each issue. The lengthy interview technique of the MJI and use of probe questions to each dilemma, whilst susceptible in terms of reliability, clearly has the potential to elicit more information than a task where participants are required to supply written responses to a single question concerning a moral issue. It may be that for the "Living even if you don't want to" item, some of the shorter written answers do not provide the rater with enough information to code the item appropriately. Whilst the authors of the measure do advocate the use of "probe" questions in circumstances where insufficient information is supplied to an item, it should be remembered that the measure is presented as one that easily be administered as a pen-to paper task in group settings, and even, amongst adults, by mail (Gibbs et al 1992:43).

In Kohlberg & Elfenbein, 1975, illustrations are given of different "stage" responses to a moral dilemma asking if a doctor should "mercy-kill" a dying patient (which on the face of it is a similar moral issue to "living even if you don't want to"). This is a small excerpt from the "Stage 2" response of Jim aged 13: *"If she requests it, it's really up to her"* (Kohlberg & Elfenbein, 1975:624); the following is taken from the "Stage 5" response of Jim aged 22: "... *If it's her own choice, I think there are certain rights and privileges that go along with being a human being...* " (Kohlberg & Elfenbein, 1975:625). These are snippets taken out of the context of much lengthier replies, and out of context, they are perhaps difficult to code reliably. It is only by examining the complete protocols supplied to this Kohlberg dilemma that one is clearly able to identify maturational differences in the types of responses given (Kohlberg & Elfenbein 1975:624-625). It is possible then, that for this type of issue a single probe question may not always be sufficient to elicit "scorable" moral reasoning data, particularly amongst younger participants.

It is possible that for adolescents the item "Living even if you don't want to" is one that is perceived as less of a "moral issue" than some of the other issues under consideration in the SRM-SF. Some indication that this might have been the case came from the comparisons of the "importance" ratings of the items that were presented at the end of Chapter 2. As reported in the overview to the measure, Gibbs et al (1992) claimed that the issues put forward in the questionnaire represent commonly held values. Gregg et al (1994) corroborated this claim with their US based sample, with the majority of subjects rating all items as "important" or "very important". However, It was the "living even if you don't want to" question that received the lowest (88%) importance rating, with all other items achieving more than 90% consensus on importance. In the current study, when importance ratings were compared, care was taken to identify those who "unequivocally" considered the issue to be "important", thus the overall findings regarding "importance" ratings were somewhat lower than in the Gregg, Gibbs & Basinger (1994) study. Still, even with this modification it was found that the majority of participants rated SRM-SF issues as "unequivocally" important. However, as with the Gregg et al (1994) study "Living even if you don't want to" had the lowest importance rating of all the items (overall 61.5% rated this item as unequivocally important – compared to a 97% consensus for the highest rated item "Saving the life of a friend").

Berkowitz et al (1995) point out that self-harm – of which suicide is perhaps the most obvious form - occupies an unusual prescriptive status in Western culture. On the one hand, it is considered morally wrong, yet on the other it is frequently argued that self-harm should not be subject to social control (Berkowitz et al 1995:202). Berkowitz et al (1995) note that whilst there is a considerable body of research that has investigated how adolescents make psychological meaning of self-harm, and in particular suicidal behaviour, little is known about how adolescents make moral meaning of self-harm (Berkowitz et al, 1995:215). They cite evidence to suggest that adolescents may be more likely to view acts of self-harm as "prudential events" rather than as moral events, and that "prudential events" are perceived as "less wrong" and "less important" than moral acts impacting on others. Clearly the response "it's up to them" that was frequently elicited to the "Living if you don't want to" in the current study would seem to indicate that for a number of the adolescents this issue may be perceived as a matter of personal choice rather than as a moral act impacting on others. Furthermore,

given that in the past few years there have been a number of highly publicised "Right to die" cases presented in the media, it is perhaps hardly surprising that this question resulted in a more equivocal response set than other issues in the SRM-SF. Given the prevalence of self-harm and risk-taking amongst adolescents, Berkowitz et al cite the need for further research into this area as a matter of urgency (Berkowitz et al 1995). Researchers interested in pursuing this area of investigation might wish to explore ways that the "Living even if you don't want to" item could be elaborated on in order to elicit information that provides more insight into the adolescent perspective.

When it came to the item "How important is it for judges to send people who break the law to jail?" a number of unscorable responses took the form of "it depends" or "it depends what they've done". On the face of it, the wording of this item makes it conceptually less straightforward to answer than some of the other items in the questionnaire. The wording may be particularly problematic for an adolescent sample – given that a number of the law-breaking activities commonly engaged in by adolescents (for example status crimes such as smoking and drinking under-age) do not usually result in custodial sentences.

The "judges sending people to jail" item seems to be qualitatively rather different in terms of "perspective taking" than some of the other items in the SRM-SF – for example "How important is it to keep a promise to a friend?". It would seem to be rather easier for a young person to imagine themselves in the position of the person keeping a promise to a friend than in the position of a judge sentencing a criminal. Furthermore, given that some of the young people answering this question were young offenders, it is perhaps unrealistic to expect that they would be willing or able to adopt the perspective of an individual giving out a jail sentence. As Krebs et al (1997) note, there are several "performance" factors that have been found to impact on moral reasoning, these include variations in "self "rather than "other" as the focus of the moral decision and variability in the consequences for the protagonist and recipient of the moral act. In this case, for the young offenders at least, it seems plausible that "self" is more likely to be viewed as the potential object of the punishment rather than "other". In a similar vein, it was interesting to note that on the item "Keeping a promise to a child", participants – including young offenders – tended to perform rather better than

on the other items in the questionnaire. For this issue, it may have been the case that the young people took the perspective of the recipient - the child – rather than of the person fulfilling the obligation. Amongst adults, perspective taking in terms of identifying with the protagonist or the recipient on both the "Judges sending people to jail" and "keeping a promise to children" might have been rather different than amongst an adolescent sample. As this is not an issue that seems to have been addressed in the SRM-SF literature, it is one that would seem worthy of exploration in future research.

4.24 Issues Relating to Aspects Within Stages:

In line with Kohlbergian theory the key emphasis of the SRM-SF is on the global structure of the moral reasoning justifications made rather than the on the content of decisions made about the various issues under consideration. Nevertheless, through the inclusion of "Aspects" within moral stages the SRM-SF should allow for the exploration of responses with reference to moral orientations within developmental levels. Indeed, during the initial "Self-training" phase of SRM-SF familiarisation one is advised to code the practice exercises according to the Aspects within stages. However, from perusal of the manual, particularly when referring to the Transitional Stages, it would appear that the "Aspects" may need further clarification if one is to gain maximum insight into differing moral orientations from SRM-SF data.

It was noted when carrying out the categorical analyses that there was some difficulty interpreting the findings relating to the Transitional Stages and that attempting to resort to the Aspects within stages would not resolve this issue. The way in which aspects from adjacent major stages had been combined to form the content categories in the Transitional Stages did not always appear to be entirely theoretically coherent. Furthermore, there appeared to be some variation in the content categories according to the moral issue under consideration.

Up until now, for the majority of researchers who have used the SRM-SF, potential problems associated with the Transitional Stages and Aspects within stages have not been issues – in line with the guidelines in the manual, findings are presented with reference to summary stage scores, most usually as an average over all questionnaire items. However, given that in the current study, findings

differed according to the technique of analysis that was adopted, (for example, whilst there were no overall differences in male and female non-offenders in summary means, clear gender differences emerged once the data were analysed using categorical techniques) and the identified need in the forensic research literature to represent intervention outcomes using indices more sensitive to change than summary means, it is apparent that these issues are important ones to address in future research.

Whilst it is almost undoubtedly the case that Gibbs et al (1992) have derived the "facets" or "Aspects" that define the moral stages on the basis of their considerable experience with other measures of moral reasoning, this is not explicitly stated in the SRM-SF manual. Furthermore, although it seems plausible that the protocols provided as scoring exemplars are representative of those most commonly provided in the construction phase of the measure, no specific information is given by Gibbs et al (1992) on this point. From the manual, there is no indication as to whether protocols represent more "typical" responses to an issue than others. The moral reasoning research goes back several decades, with MJI protocols being initially derived from US based males in the 1950s. It goes without saying that cultural and historical changes may have considerable impact on the nature of discourse surrounding moral issues. In the current study, coding the protocols was carried out with strict reference to the SRM-SF manual - no attempt was made to analyse the data in qualitative terms (the size of data set in the current study would have posed considerable practical difficulties to this type of analysis). However, given the apparent problems in classifying some of the responses made to some of the items, it may be the case that by adopting a "Grounded Theory" approach (Strauss & Corbin 1998) to SRM-SF data one might elicit moral themes that better represent the moral discourse of UK based adolescents in the 21st Century.

4.25 Issues Relating to Age of Participants:

It seems from some studies using the SRM-SF that more attention needs to be paid to the ages of the participants in the sample. Bearing in mind that the SRM-SF is a developmental measure it was somewhat surprising to discover during the literature review that in some studies (e.g. Palmer & Hollin, 1998) participants of a heterogeneous age range were treated as a homogenous group. It should be noted that this strategy was also adopted in the meta-analysis carried out by Jaffee & Hyde

(2000) that attempted to resolve conflicts in the literature regarding gender differences in patterns of moral reasoning. In their meta-analysis all "young people" aged between 11 and 19 were categorised as "adolescents" (Jaffee and Hyde 2000). Furthermore, the ages of some of those classed as adolescents were estimated – in some cases the midpoint of the sample age range was used. Therefore one cannot be clear from the meta-analysis what the actual age distribution of the adolescent sample was. Neither was it possible to tell whether there were disproportionately greater numbers of older females than younger males in the sample, or vice versa. So whilst Jaffee and Hyde noted that the magnitude of gender differences appeared to be greatest during adolescence it is not possible to tell from their study when and where these differences manifested themselves.

Taking account of age disparities would seem to be particularly important when investigating moral reasoning amongst an adolescent sample. In the current study, once the sample was categorised according to age groups, information was revealed that was not identifiable from the sample as a whole. For example, when comparing overall between-group differences amongst male and female non-offenders no differences were found in overall means. However, once the sample was partitioned according to age bands it was found that younger female non-offenders had significantly higher means than the male non-offenders. Furthermore, whilst in the overall sample male offenders had significantly higher means than the offenders, in the youngest age group category there were no significant differences between the offender and male-non offender means. Differences between these two groups only emerged in the second age group band. Whilst the data were cross-sectional, the categorisation of participants according to age enabled one to make comparisons across age bands, these comparisons indicated that as a group the male non-offenders were showing more dramatic shifts in terms of moral reasoning means across age bands than either of the other two groups. Clearly, information of this nature is potentially extremely valuable when seeking to gain insight into normative patterns of moral development amongst adolescents; furthermore, adolescence is a time where marked maturational changes may take place over a span of months rather than years. It seems rather curious, then, that some researchers using the SRM-SF have apparently paid so little attention to potential age-related confounds within their sample. For example, in the Jaffee and Hyde (2000) meta-analysis where young people from 11-19 years old were classified as "adolescents", differences

in "justice" and "care" reasoning as evidenced in moral reasoning scores amongst older adolescents may have been partially cancelled out due to known gender differences in maturational rates amongst the younger participants. Moreover, when considering moral reasoning differences between an offending and a non-offending adolescent sample, taking account of relative chronological age would seem to be paramount, particularly when considering clinical interventions with young people.

4.26 Ignoring "Flags" to Variability Present in Research Findings:

In the previous sections, some attention has been paid to the way in which certain aspects of the scoring and analysis procedure of the SRM-SF may contribute to the masking of heterogeneity in stage use within the sample. However, it also appears from some studies that when evidence of heterogeneity in stage use is found within samples, there is a tendency to overlook this type of information. For example, a review of the literature found that although in some studies using the SRM-SF Standard Deviations were presented along with summary means, typically researchers tended to mention variation around the mean "in passing". In the current study, it was found that whilst there was little difference in overall means of males and females, the standard deviation was higher amongst male non-offenders. As not all moral reasoning studies present their findings relating to standard deviations around the mean, there were difficulties in establishing whether this finding was representative of previous work. Nevertheless, there seemed to be no indication from the available research that the standard deviations were aberrant.

Clearly, information that one can glean from the standard deviations around moral reasoning means is potentially useful in a variety of ways. Between-group differences in magnitudes of standard deviations may add weight to, or qualify findings relating to group means. For example, as noted above, in the current study standard deviations around overall means were higher amongst male non-offenders than amongst the other two groups. When one takes this finding in conjunction with the finding that between-group differences between offenders and male non-offenders in overall means were statistically significant, added weight is given to a claim that in a mixed-age sample offenders are demonstrating lesser evidence of developmental progression in moral reasoning than male non-offenders. The finding that male and female non-offenders in the oldest age group category had

similar means (male non-offender means = 278.74, female non-offender means = 272.41) but somewhat different standard deviations (male non-offenders SD = 41.14, female non-offenders 29.31) provided one with the preliminary indication that males were demonstrating greater heterogeneity in mature stage use that was later corroborated through the use of categorical analyses. It would appear, then, in the SRM-SF research, that when presenting findings relating to summary means, greater attention needs to be paid to sources of variability as evidenced through standard deviations.

4.27 Patterns of Intra-Individual Variation

The current study appears to be the first to make a systematic attempt to investigate levels of intra-individual variation in stage use with reference to SRM-SF data. Therefore one cannot tell whether the findings in the current study were idiosyncratic. From the analyses that were presented in Chapter 3, it was apparent that participants were demonstrating relative intra-individual variation in stage use across SRM-SF items, there were problems, though, when it came to trying to pin-point the sources of this variation. A review of work carried out in this area using alternative moral reasoning measures such as the DIT and the MJI revealed that aside from gender differences two factors figured prominently in the literature: "external" situational factors potentially influential on moral reasoning performance, and variations in patterns of "consolidation" and "transition" associated with "internal" mechanisms of developmental change. However, on further investigation, neither model seemed adequate in itself to account for the patterns of variation observed in the current study.

As reported earlier, difficulties in applying the "consolidation and transition" model to SRM-SF data may stem in part from the scoring procedure of the measure – assigning summary scores to each item, rather than including all raw stage-typed data elicited by participants. A further study that compared differences in patterns of stage distribution obtained with standard SRM-SF data with those obtained when including all stage typed information provided to items might be able to shed further light on this issue.

Whilst it seemed clear that overall patterns of intra-individual variation may have partially resulted from situational factors relating to the nature of the issue under consideration, certain variations in task conditions did not seem to impact on levels of intra-individual variability. Furthermore the finding that male non-offenders demonstrated significantly higher intra-individual variation than the other two groups poses problems when seeking to interpret the findings purely with reference to a situational framework. Male and female non-offenders had all performed the task under the same environmental conditions.

As reported in the introduction to this thesis, researchers investigating sources of variability in moral reasoning have turned their attention to a wide range of potentially mediating variables. Indeed, as Krebs et al (1997) point out, the study of performance factors impacting on moral reasoning is potentially never-ending (Krebs et al 1997:135). Typically, though, in the literature relating to intra-individual variability in moral reasoning, one finds that the research is somewhat polarised, with studies tending to focus either on external situational factors or on internal person-related characteristics as mediators of moral reasoning. Rather less attention in the research has been paid to the investigation of the potential interaction between these types of processes.
For example, Narvaez, et al (1999) note that in the study of moral development two models are typically put forward to explain change: "cultural socialisation" and "internal cognitive construction of moral meaning" (Narvaez et al, 1999:478). However, in the literature these models are frequently treated as "rival explanations" rather than as mutually interactive (Narvaez et al 1999:479).

Those studies that have sought to investigate the interaction between person-related characteristics and situational factors as mediating influences on moral reasoning have typically compared gender differences in stage use across so-called "real life" and "hypothetical dilemmas". The results of these studies have been mixed. For example, whilst a number of studies demonstrate that participants' moral reasoning varies across types of dilemma (e.g. Haviv & Leman, 2002; Wark & Krebs, 2000; Wark & Krebs, 1996; Jadack et al, 1995; Carpendale & Krebs, 1992; and see review by Krebs et al 1997), the results regarding the impact of gender are conflicting. Furthermore, due to variations in experimental design and potential confounds associated with some of the studies, the findings are

sometimes hard to interpret (further issues relating to some of this previous work will be discussed when considering the implications of the gender differences observed in the current study).

It would seem plausible that the patterns of intra-individual variability observed in this sample may be the result of a complex interaction between age, gender, offence status, environmental conditions and nature of the issue under consideration. For example, whilst the male and female non-offenders carried out the task under ostensibly the same conditions – as a classroom exercise – it may be the case that the task condition has the potential to elicit more intra-individual variability amongst males than amongst females. It was noted in Chapter 3 that in a study carried out by Galotti et al (1991) where males and females were asked to carry out an open ended essay task giving examples of personally relevant moral issues females tended to respond more thoroughly to the task than males. In the UK, females are now out-performing males in terms of academic performance at all levels of secondary education, it may be the case that the written task was more onerous for some of the males than it was for the females. On the other hand, given that moral reasoning protocols in measures such as the MJI were initially generated on the basis of male moral justifications, the SRM-SF may simply be more sensitive to intra-individual variation amongst males than it is amongst females. Were this to be the case, then the finding that male non-offenders demonstrated greater intra-individual variation than the (male) offenders might be attributable to differential patterns of developmental change in the two samples - mirroring the findings obtained when comparing between-group differences in the standard deviations of the participant groups.

Aside from those longitudinal studies exploring patterns of consolidation and transition amongst young people, much of the more recent work exploring internal structural flexibility in moral judgement within males and females has been carried out amongst adults – frequently university students - based in the US or in Canada. Less attention has been paid in the research to exploring potential sources of intra-individual variation amongst adolescents. Furthermore, whilst there are a host of studies that have investigated potential mediators of inter-individual differences in moral reasoning amongst offenders, little interest has been shown in the research in exploring patterns of intra-individual variation in moral reasoning amongst this participant group. Bearing in mind the

recent concerns expressed in the forensic literature regarding the efficacy of summary indices to identify clinical need it seems clear that further research in this area is required.

4.28 Observations Regarding the Offending Sample.

In line with a now well-established body of research, the initial analyses comparing overall SRM-SF means of offenders and male non-offenders revealed significant between-group differences, with non-offenders demonstrating higher means than non-offenders. It was of course useful to be able to corroborate previous work in this area, however, the finding in itself does little in terms of providing an original contribution to the field and is limited in terms of its clinical applicability. The findings resulting from the subsequent analyses that were carried out are, however, potentially more interesting in terms of novel information that was provided.

Researchers using the SRM-SF to compare the moral reasoning of offenders and non-offenders have tended to attribute the poorer performance of the offending group to a "pervasive cognitive structural immaturity" (e.g. Gregg et al 1994; Palmer & Hollin, 1998). However, in the current study, whilst offenders were overall apparently globally "at" a lower level of moral reasoning than their non-offending peers, the results of the subsequent analyses would seem to call this type of interpretation into question. From the results of the comparisons of the SRM-SF summary means of offenders and male non-offenders in Age Group 1 presented in Chapter 2, one was not able to evidence a global structural deficit in the offending group. Although the mean scores were higher amongst the male non-offenders, between-group differences were not statistically significant. In Chapter 3, once the data were analysed in categorical terms it was found that whilst overall non-offenders were more consistent in their use of "mature" stages, a substantial majority of the offending sample were providing "Stage 3" justifications to at least one item in the SRM-SF.

In the Discussion in Chapter 2, it was acknowledged that the results pertaining to the younger males may have been idiosyncratic to the current study. The sample size available for comparison was rather small, it may have been the case that had one made between-group comparisons amongst a larger data set significant between-group differences would have emerged. Furthermore, it was noted

that whereas in some studies (e.g. Gregg et al, 1994) efforts were made to "clean" the non-offending data set by eliminating school students perceived by teachers as exhibiting conduct disorders, no attempt in the current study was made to clean the data set in this way. However, it was also pointed out that a certain amount of data "cleansing" of the non-offending sample would probably have occurred by default, with those pupils excluded from school for challenging behaviour being naturally eliminated from the sample. Moreover, it was posited that the younger offenders may have represented an extreme point of the spectrum in terms of persistent anti-social behaviour, having been convicted of a criminal offence at a very young age. For example, if one refers to recently published statistics comparing patterns of crime during the adolescent years, one finds that in 1998 27,300 young people were found guilty or cautioned for a criminal offence in the 12-14 age group, this rose sharply to 67,200 in the 15-17 age group, and then levelled off somewhat in the 18-20 age group 77,500 (Coleman & Schofield, 2001:87).

As was pointed out earlier, previous studies using the SRM-SF to compare moral reasoning levels of young offenders and non-offenders have used samples with rather broad age ranges. Furthermore, the average age of both offending and non-offending participants in those studies was somewhat older than the average age of the participants in the current study. As far as the author is aware, this was the first study to date that sought to investigate differences in SRM-SF scores between younger offenders and non-offenders. In order to clarify the position one would need to carry further comparative studies amongst male offenders and non-offenders in this age group. Further research of this nature would seem to be essential, as if the findings observed in the current study are replicated, then this may present problems for those seeking to use the SRM-SF as an assessment tool with younger offenders.

It may be the case that although the differences in mean scores between the younger (male) offenders and male non-offenders are not statistically significant, they do represent a critical difference in terms of criminogenic risk. As Friendship et al (2003b) point out in their discussion of the assessment of clinical interventions, differences in effect size ignore the relevance of the level of functioning amongst participant groups (Friendship et al 2003b:120). The male non-offender means were higher

than offenders by some 10 points (male non-offenders = 228.84, offenders = 218.41). This difference may seem minimal numerically, however, it may represent an important one in terms of identifying clinical need amongst the offending group.

Clearly, the difficulties associated with the GMSS scale that were identified in Chapter 3 – most notably as a result of intra-individual and inter-individual variability in stage use - preclude one from using this scale as a reliable index of the moral reasoning levels of participants. Nevertheless it is perhaps interesting to note that the mean scores of the younger offenders and male non-offenders correspond to differing levels of the GMSS scale. The mean score of the offenders in Age Group 1 (218.84) corresponds to global GMSS level "Stage 2" whereas the mean score of the non-offenders (228.41) corresponds to global GMSS level "Stage 2(3)". In theoretical terms, then, if a future study replicated the findings amongst a "structurally consistent" sample the higher score of the non-offenders mean scores would represent the first signs of the morally desirable shift from Stage 2 reasoning towards Stage 3 reasoning (if one adopts the Gibbs et al (1984), view that Stage 3 acts as a "cognitive buffer" against anti social behaviour).

Given that the first indications that there is a shift towards Stage 3 is manifested in the production of "Transition Stage 2-3" moral justifications, one might seek to explore patterns of stage amongst the younger males with reference to the frequencies of Stage 2 and Transition Stage 2-3 mix across the eleven items of the questionnaire. However, as was pointed out in Chapter 3, problems were encountered when seeking to interpret the data set in this way. Clearly, the overall level of variability in range of scores that was problematic when seeking to interpret the findings with reference to GMSS ratings is also an issue when seeking to characterise the data with reference to potential proportions of "adjacent stage mixes". It was found that in the current study, very few of the participants restricted their stage use to adjacent levels only. Furthermore, as not many participants in the current study provided scorable responses to all eleven items in the questionnaire, one could hardly claim that profiling the data set in this way was representative of the participant sample as a whole. Finally, there is, in any case, as far as the author is aware, no information currently available in the SRM-SF regarding typical response sets associated with the moral developmental levels.

However, for those interested in using the SRM-SF with younger males in future research it may be of interest to note that if one were to "translate" the mean scores of the younger offenders and male non-offenders into proportions of Stage 2 and Stage 2-3 mix across eleven items, the differences in the mean scores would result in rather differing profiles for the two groups. In terms of a "Stage 2" and "Transition Stage 2-3" mix the mean score of the offenders approximates to seven "Stage 2" responses and four "Transition Stage 2-3" responses (this mix results in a score of 218.18) whereas the mean score of the male non-offenders approximates to five "Stage 2" responses and six "Transition Stage 2-3" responses (this mix results in a score of 227.27). In an "ideal case" scenario, then, where there is evidence of structural consistency amongst participants providing scorable responses to all SRM-SF items, whilst the younger offenders are showing majority "Stage 2" and minority "Transition Stage 2-3" use across the SRM-SF, the reverse is true of the non-offenders. Thus a minimal difference in SRM-SF means may represent a very real difference in terms of moral reasoning developmental levels.

As was noted earlier, the representation of the moral reasoning evidenced to individual SRM-SF items through the use of one summary score alone may mask important information regarding raw stage typed data being elicited through multiple justifications to issues. This may be of particular importance when seeking to identify differences between younger males. As reported in the Discussion in Chapter 2, typically, in the research when comparing the moral reasoning of offending and non-offending adolescents one finds that offenders are comparatively immature in relation to non-offending peers, and that there is greater evidence of "Stage 3" use amongst the non-offenders. It was pointed out that whereas it may be relatively straightforward to make a distinction between proportions of "Stage 2" use and "Stage 3" use amongst an adolescent sample with a broad age-range, or amongst older adolescents; in theoretical terms one might expect to encounter difficulties when seeking to make the same kind of distinction amongst the younger males, as "Stage 2" represents the norm for this age group. However, it may have been the case that had one made reference to <u>all</u> the types of moral information being provided to each item, one might have been able to better discriminate between the younger offenders and male non-offenders. For example, one might have found through scrutinising multiple justifications provided by the younger males that whilst summary

scores attained for items were similar in terms of being coded as "Stage 2" or "Transition Stage 2-3", performance levels in terms of consistent production of these stages within items differed between groups. Bearing in mind that one of the reasons offered as a possible explanation of the apparent failure to apply the "consolidation and transition" model to the SRM-SF scores in the current study was paucity of raw stage type data, this would seem to be an area worthy of further investigation in the research.

A further explanation of the failure to establish global differences between the younger males offered in Chapter 2 was that the younger offenders may not be very different in terms of overall moral reasoning but may be manifesting specific deficits that are in themselves risk factors for the development of moral reasoning. The one issue where clear differences did emerge between the younger offenders and male non-offenders was on the item "Obeying the law". This was of course the one area where there were clear behavioural differences between the two groups. It was posited that moral reasoning disparities amongst younger males in this sub-domain had the potential for considerable "knock-on" effects resulting in global divergence in moral reasoning in later years. For example it was argued that continued involvement in persistent offending activity was likely to "reinforce" the "Stage 2" reasoning already present and was not conducive to a normative developmental shift towards Stage 3 "Mutual and Prosocial" reasoning. It was pointed out that criminal activity is more easily justifiable to the young person if they focus on "pragmatic" considerations rather than taking the perspective of the victim; it was also noted that continued involvement in offending behaviour was likely to have a negative impact on those "normal" social relationships considered essential by Kohlbergian theorists for the development of global moral competence.

The finding that younger offenders and non-offenders differed on the item "How important is it to obey the law?" is potentially interesting, nevertheless this item represents only one out of the eleven issues under consideration in the SRM-SF. Furthermore, the item is very broad in terms of its scope, and is not informative in terms of addressing those specific types of law breaking activities commonly engaged in by young people. The one item that does address specific aspects of law breaking

behaviour – "How important is it not to take things that belong to other people"- addresses "theft" in a rather indirect way and failed to discriminate between the younger offenders and male non-offenders.

When scoring the item "How important is it to obey the law?" it was noticed that common justifications elicited from the young-offenders took the form of "Otherwise you will / might get into trouble". This type of response is deemed to be characteristic of the immature Stage 1-2 or Stage 2 reasoning more prevalent amongst young offenders than non-offenders. However, moral justifications of this nature tell one little in terms of relating moral reasoning to moral action. The young offenders as a group had all broken the law, and they had all "got into trouble" as a result. One cannot tell from this response whether the young person is addressing this issue retrospectively with reference to the consequences of previous law-breaking behaviour or in terms of considering current / future conduct. It should also be noted that amongst British adolescents "The law" is frequently used as a slang referent for "the police"; therefore some of the young people may have been answering this question in terms of obeying figures of authority rather than in terms of conforming to specific social norms.

If, as the result of future research it transpires that younger (male) offenders and non-offenders are little different in terms of global moral reasoning, but do manifest identifiable differences when it comes to the sub-domain relating to offending behaviour, then clearly this is an area that should be explored in more depth. It should be remembered that the types of "offending behaviour" most typically engaged in by adolescents encompass a rather restricted range of criminal acts: the majority of "young offenders" do not commit the majority of the offences that are on the statute books. In future studies, researchers investigating moral reasoning amongst younger offenders might consider "tailoring" the SRM-SF in order to address those specific criminal activities most commonly engaged in by young people (e.g. status crimes, acquisitive offences, criminal damage, car-crime and drug offences).

Whilst the results of the comparisons of the overall means of the offenders and non-offenders corroborated previous research in the field indicating that offenders were morally immature in relation

to their non-offending peers, the results of the subsequent analyses were less conclusive in this respect. In particular, the findings from the categorical analyses presented in Chapter 3 that the majority of offenders were apparently able to spontaneously produce "Stage 3" justifications to at least one item in the SRM-SF raised issues regarding the pervasiveness of moral reasoning cognitive deficits amongst this population. The findings obtained from the categorical analyses can be considered to be "new" in the sense that they arose from a novel approach to tackling SRM-SF data; however, they can hardly be said to be "news" to researchers in the field. As far back as 1980 both Blasi and Jurkovic expressed caution about interpreting moral reasoning differences amongst offenders and non-offenders with reference to underlying global structural deficits.

It may of course have been the case that the findings relating to the spontaneous production of "Stage 3" reasoning were idiosyncratic to the current study. However, it should be recalled that the offending sample was drawn from a wide variety of different settings spanning the length and breadth of the UK. Evidence of Stage 3 reasoning was provided by at least one young offender in every participant setting. Furthermore, there was no indication from the initial comparisons of overall means that the offenders in this sample were performing at a higher level of moral maturity than has been evidenced in previous research. Indeed, the reverse may said to be the case. The average age of the participants in the current study was younger than in other comparable studies using the SRM-SF. The overall means of the offenders in the current study were actually somewhat lower than those reported by other researchers (for example Gregg et al, 1994 report mean SRM-SF scores of 243.10 for male adolescent offenders in their study).

In the literature, offenders are most frequently referred to in terms of their limitations and deficiencies compared to non-offending peers rather than in terms of capabilities and competencies (Carroll et al 2001). Indeed, as reported earlier, researchers using the SRM-SF to compare the moral reasoning of offending and non-offending adolescents frequently interpret their findings regarding the poorer performance of offenders on the measure to underlying pervasive structural deficits (e.g. Gregg et al, 1994, Palmer & Hollin, 1998). This emphasis in the literature is of course hardly surprising, if one considers the need to identify underlying criminogenic risk factors, and the assumption of structural

consistency in moral reasoning that underpins Kohlbergian theory. However, for practitioner and young offender alike, this stress on global inadequacies in the research does little in promoting motivation to change. In the moral reasoning research the convention to present moral reasoning findings with reference to summary means reinforces the view that "deficiencies" amongst young offenders are global and fixed, rather than specific areas of need that are responsive to remedial intervention.

For the academic researcher establishing that there are apparent global deficiencies in the moral reasoning of offenders compared to non-offending peers and that these deficiencies manifest themselves amongst diverse age ranges in disparate cultural settings is of considerable importance. Ascertaining that there are global trends of this nature contributes to our general understanding of psychological factors implicated in offending behaviour amongst young people, and points the way towards future research investigating the potential underlying causal influences of this offending behaviour. For the criminal-justice practitioner, however, whilst it is clearly of use to know that moral reasoning "deficits" represent a criminogenic risk factor, academic research findings may contribute little of clinical relevance to intervention work being undertaken in the field. For the practitioner, the key emphasis is not on understanding underlying global causal factors relevant to the general offending domain, but on promoting socially desirable behavioural shifts amongst individuals.

There are now a number of studies using the SRM-SF that have identified those global differences in moral reasoning amongst offenders and non-offenders well documented in research using alternative measures of moral reasoning. It was clearly necessary to carry out this type of research for validation purposes and in order to demonstrate that the SRM-SF reliably discriminates between offenders and non-offenders amongst diverse populations. However, now that the necessary ground-work has been done in validating the measure, attention should turn towards ways in which academic research using the SRM-SF can be geared towards practical work in the field. It is of course much easier for an academic researcher to provide evidence of global moral deficiency amongst a large sample of young offenders than it is for a practitioner to demonstrate that there is a desirable moral shift at post-test amongst a small number of young people participating in a group-work intervention. Earlier in this

discussion, concerns were expressed that SRM-SF summary means may not in themselves be sufficiently sensitive to detect the slight or at best moderate global shifts in moral reasoning that one could realistically expect to find at outcome of a short-term programme of intervention. Clearly lack of sensitivity of the measure limits its clinical applicability.

It is extremely disheartening for practitioner and young person alike when perceived meaningful improvement in a specific domain as the result of a moral reasoning intervention is not reflected by a corresponding change in global scores at post-test. Furthermore, lack of "statistically significant" change in overall means may mask within-group and within-domain improvements that are of some considerable clinical relevance. However, whilst research findings relating to overall SRM-SF means may be limited in terms of their practical applicability, there are other aspects of SRM-SF research that are potentially more useful in this respect.

Focusing on specific content categories within the global moral reasoning domain may be one way in which moral reasoning theory can make a more valuable practical contribution to the field. From the evidence provided to date in the SRM-SF literature it is apparent that offenders perform less well than non-offenders on the SRM-SF. However, from those few studies that have compared patterns of moral reasoning across individual SRM-SF items there is some indication that young offenders may be more disparate than non-offenders in some areas of the moral reasoning domain than in others. From the evidence of the current study it was found that offenders' moral reasoning was lower on the socio-moral values relating to offending behaviour than on other sociomoral values in the SRM-SF; similar disparities in performance across moral values were also observed by Palmer & Hollin (1998) and Gregg et al (1994). If as the result of future research it is established that this represents a general trend, then this may have significant implications for the effective targeting of clinical interventions. For example, if one finds that amongst an offending sample that "deficits" in moral reasoning are most notable when it comes to "Justice" issues, but that the offenders are demonstrating greater evidence of moral maturity when it comes to "relationship" issues, then the focus of the moral reasoning intervention may be on promoting the generalisation of moral competence from specific areas to the global domain. On the other hand, if amongst an offending sample one finds that there is

no evidence of reasoning above Stage 2 to any of the issues in the SRM-SF then the focus of the intervention will be on the best way to stimulate and initial morally desirable shift. Knowledge that some content areas may be more amenable than others in this respect is clearly of value. For example, one wish to establish evidence of moral maturity within the "relationship" sub-domain before attempting to stimulate shifts in a potentially more challenging "legal justice" domain. It is worth noting that the poorer performance on the items relating to offending behaviour was not merely restricted to the young offenders in this sample. The legal justice domain may be one that is more susceptible to "downward press" on moral reasoning performance than other aspects of the domain.

Whilst it is clearly necessary to challenge distorted or morally "immature" thinking amongst offenders participating in a remedial intervention, highlighting those areas where offenders are doing "as well as" their non-offending peers may be equally important. As a result of their experiences within the criminal justice system many of the young offenders will have been made all too aware of their shortcomings compared to same age non-offending peers. If practitioners and /or the young people tend to view these shortcomings as global and stable attributes then this presents challenges when it comes to motivating young people participating in a clinical intervention. As Gibbs (2000) points out Kohlberg's theory is not in itself enough to motivate young offenders to change. There is of course a considerable difference between demonstrating that an individual is not performing as well as other people and demonstrating that an individual is not performing at their own optimal level. Furthermore, focusing solely on the ways in which young offenders differ from their non-offending peers may actually be counter-productive in motivational terms, especially if one adopts the Emler and Reicher (1995) view that delinquency is a form of self-presentation and that some young offenders may seek to enhance their "delinquent reputations" (Emler & Reicher, 1995:7).

As was pointed out earlier, whilst offending amongst adolescents is perceived to be widespread, the offending behaviour typically engaged in by young people is restricted to a relatively narrow range of criminal acts. The majority of offenders commit a minority of offences. As Emler & Reicher (1995) note whilst "non-conformity" tends to attract the limelight in the research literature as soon it becomes a social problem, the absence of delinquent behaviour is of as much interest psychologically as its

presence (Emler & Reicher 1995:8). Jamieson et al (1999) point out that in the research one finds a division between studies that focus on the causes of offending behaviour and those that concentrate on factors influential in the cessation of offending behaviour. Their recent study amongst Scottish adolescents attempted to bridge this gap by investigating choices made by young people regarding "resistance to", "desistance from" and "persistence in" offending. They suggested that as the bulk of research in this field stems from the US there was a clear need for further empirical work in this area in a UK context (Jamieson et al 1999:1-3).

Gaining further insight into factors underlying "resistance to and "desistance from" offending is clearly as important as understanding why some young people persist in engaging in criminal activity. However, further understanding in this field may not give us direct insight into young people are motivated to conform to some social norms and not others. The main emphasis in the offending literature has been on understanding underlying influences relevant to anti-social behaviour. Far less attention has been paid in the research on factors that promote conformity and/or pro-social behaviour amongst a young offending population. In the current study, when reading through the responses given to the item "How important is to obey the law?" the author gained the impression that the majority of young people had responded to the question with reference to potential negative consequences of breaking the law rather than with references to the positive benefits of law-abiding behaviour. Whilst there was an overall tendency for young people to respond to SRM-SF items with reference to negative consequences of transgressing the norm rather than to potential positive outcomes of conforming, this seemed to be less the case for some items than others. For example, whilst it appeared that some responses to the item "How important is it to keep a promise to a friend" focused on potential consequences if the promise was broken (e.g. "otherwise your friend might get upset"), a number made reference to the benefits of keeping the promise (e.g. "It shows your friend you can be trusted"). It was noted earlier that responses provided by young offenders to the item "How important is it to obey the law" frequently took the form of "Otherwise you will/might get into trouble". Whilst these responses are indicative of the comparative moral "immaturity" associated with group, this type of response can hardly be considered to be informative in explaining individual behavioural choices young people had made regarding resistance to, desistance from or persistence in

criminal activity, and / or why offenders might consider it important to obey some laws and not others. Furthermore, it should be noted that the type of response cited above was relatively common amongst non-offenders as well as offenders. Future SRM-SF research might seek to explore between-group and between-item differences in "transgression or conformity" issue focus: for example by exploring whether young offenders are more likely to focus on the consequences of transgressing a norm than non-offenders, and/or whether some issues are more likely to elicit "transgression focused" responses than others. A possible extension of this type of research would be to include additional issues in the SRM-SF that relate more directly to pro-social activity (for example with reference to helping, sharing and donating activities). One acknowledges that there may not necessarily be a linear relationship between increase in pro-social activity and reduction in anti-social behaviour; however, there would seem to be a need to shift the emphasis in the offending-behaviour research from focusing solely on the negative attributes associated with this group, towards gaining a greater understanding about positive qualities pre-existent in this population that one would seek to foster and nurture.

4.29 Observations Regarding Gender Differences

Some of the most interesting findings to emerge from the current study concerned gender differences in stage use amongst the non-offending adolescents. As noted in the introduction to the measure, the majority of previous work using the SRM-SF has failed to uncover evidence that the measure is unfavourably biased against females. Indeed, when comparing SRM-SF means of younger male and female adolescents it has often been found that females out-perform males on the measure: when this finding has emerged, it has generally been interpreted by researchers as indicative of known gender differences in maturation rates in early adolescence. In the current study, the preliminary findings that were obtained from between-group comparisons of SRM-SF summary scores apparently corroborated previous work. When comparing overall SRM-SF means no significant differences were found between male and female non-offenders. When comparing SRM-SF means by age category, it was found that only in Age Group 1 were there any significant between-group differences, with females having higher SRM-SF means than males. However, on further investigation, it became clear that the position regarding gender differences was not as straightforward as it had initially appeared. The bulk

of the evidence concerning gender differences emerged from the "subsequent analyses" presented in Chapter 3; nevertheless, as previously discussed, there were preliminary indications from some of the analyses carried out in Chapter 2 that use of SRM-SF summary means as the sole source of reference regarding the moral reasoning of the sample may have masked useful information regarding heterogeneity in stage use. It was suggested that previous researchers using the SRM-SF had failed to attend to potential between-group differences in inter-individual variability when presenting their findings regarding gender differences.

In Chapter 3, when patterns of stage use were compared using categorical techniques, some significant gender differences did emerge. These included the finding that there were between-group differences in use of Stage 3 and Stage 4 (with females making greater use of Stage 3 and males making greater use of Stage 4). It was of course interesting to note that there were between-group differences in stage use amongst male and female non-offenders that apparently corresponded to the Gilligan (1982) hypothesis; however, what is possibly of more interest to the researcher is that this between-group difference only emerged when analysing the data using categorical techniques. As far as the author is aware, this was the first study using the SRM-SF that had sought to compare findings obtained when data were treated as "continuous points on a scale" and when data were treated as "categorical", therefore there is no way of telling whether the differences in outcome that were observed were idiosyncratic. One would need to replicate the findings in future work in order to clarify this issue. It would seem to be essential carry out further work of this nature, as if the findings are replicated, then this may raise serious questions regarding the interpretation of findings from some of the previous research.

In a recent article published by Barriga et al (2001) exploring the relationship between moral reasoning, gender and offending, a somewhat pessimistic view is taken of the contribution that moral stage theory has made to our understanding of antisocial behaviour in young people. Barriga et al (2001) claim that there are two notable problems in this respect. Firstly, they point out that although there is an established link between moral reasoning "stage delay" and antisocial behaviour, the magnitude of the relationship in many of the reported studies is rather small. Secondly, they suggest

that "moral stage" theory has had rather little to offer in the way of explaining gender differences in antisocial behaviour. Barriga et al argue that if "moral judgement stages" were the most important and most direct influence on moral behaviour, then one would expect to see corresponding gender differences evidenced in the moral judgement research. However, Barriga et al state that save for the acknowledged maturational differences in early puberty, gender differences in moral reasoning have been found to be "negligible". They back this statement up by drawing attention to Walker's (1984) large scale meta-analysis that "demonstrated" that gender accounted for a "nonsignificant 1/20 percent of the variation in moral judgement stage (Barriga et al, 2001:535).

In light of the findings observed in the current study the recent comments made by Barriga et al (2001) are particularly interesting. The main claim that is made - that moral reasoning theory has made only a limited contribution to our understanding of gender differences in antisocial behaviour - is based on the apparent failure of the moral reasoning research findings to provide strong evidence of between-group differences in the predicted directions. However, as has been noted at several points during the course of the current study, whilst academic researchers in the moral reasoning field frequently present their findings with reference to gross markers such as overall mean scores, use of such summary indices may mask valuable between-group and within-group information that is of considerable clinical relevance. It is possible, then, that by restricting their discussion to the apparent failure of the global moral reasoning research findings to contribute to our understanding of offending behaviour that Barriga et al may be at risk of "throwing the baby out with the bathwater".

The first "problem" highlighted by Barriga et al (2001) concerns the small magnitudes in effect size that are commonly reported in findings that have established a relationship between moral stage delay and antisocial behaviour. However, as was pointed out earlier in this discussion, differences in magnitude of effect do not in themselves say anything about differences in levels of cognitive functioning (e.g. Friendship et al (b)2003). Differences in developmental levels of functioning would appear to be key to understanding differences in moral reasoning between offenders and their non-offending peers. Typically, in the research, one finds that the apparent moral delay of offenders is interpreted with reference to an apparent failure to make the shift from "morally immature" Stage 2 to "morally mature" Stage 3. Between-group differences of similar magnitudes may have very different

developmental implications depending on where they "sit" on the SRM-SF scale (for example the difference between a mean score of 220 and one of 260 corresponds to a developmental shift from Stage 2 to Stage 3(2), whereas the difference between a mean score of 280 and one of 320 does not represent a change in developmental level, both scores corresponding to Global Stage 3). Conversely, increases in the magnitude of between-group differences do not necessarily result in greater disparities in moral reasoning developmental levels.

The second problem that Barriga et al (2001) cite relates to the failure of the moral reasoning research to "mirror" known gender differences in antisocial behaviour. Aside from gender differences in early adolescence, there is no evidence to indicate that females out-perform males in the moral judgement domain. However, Barriga et al (2001) are perhaps misguided in their assumption that if "moral judgement stages" were the most important influence on moral behaviour then one would find evidence of corresponding gender differences in the moral reasoning literature. Much of the existing research into gender differences in moral reasoning has focused on addressing the claim that moral reasoning measures are biased against females. Thus, until recently, the main emphasis has been on trying to establish whether females are disadvantaged on moral reasoning measures, not on when and where they may be morally "superior" to males. Indeed, this was the main focus of the Walker (1984) meta-analysis that is cited by Barriga et al (2001) as evidence of "negligible" gender disparities in moral reasoning. As was noted in Chapter 3, the finding that females can do "as well as" males on moral judgement tasks does not in itself provide evidence that moral reasoning measures are gender fair. Furthermore, where gender differences have been found on measures of moral reasoning they almost invariably correspond to the same trend. Females tend to demonstrate greater evidence of Stage 3 relative to Stage 4 reasoning than males, and also tend to demonstrate greater evidence of care oriented reasoning relative to justice oriented reasoning than males. Tangney and Dearing (2002:264) point out that they know of no studies that demonstrate that males have a significantly greater "ethic of care" than females. Much of the research in this field stems from studies comparing summary means from a typology where Stage 3 (which according to a strong version of the Gilligan, 1982 hypothesis represents a "care oriented reasoning" style of reasoning), is scored at a lower level than Stage 4 (which Gilligan, 1982 considers to be more "justice oriented in its style). To date, as far as

the author is aware, none of the researchers in the moral reasoning field have devised a typology where the position is reversed so that "care" reasoning sits above "justice" reasoning in a moral reasoning hierarchy. Therefore there seems no reason to expect that there would be global evidence from the research findings to indicate that females out perform males in moral judgement.
In the current study, it was demonstrated that merely by recoding the data so that arithmetical homogeneity was imposed on the "mature" stages, one could apparently change the outcome regarding overall gender differences on the SRM-SF. Whilst the findings from the current study may be idiosyncratic, it seems likely that issues relating to the measurement and statistical analysis of moral reasoning may be partially responsible for the failure to identify global gender differences favouring females in the moral judgement sphere. As Gilligan and Wiggins pointed out in 1987:
"If there are no sex differences in empathy and moral reasoning, why are there sex differences in moral and immoral behavior? Either there is a problem in the way that empathy and moral reasoning are being measured, or the role of empathy and cognition in moral development has been overstated" (Gilligan & Wiggins, 1987:279).

For Barriga et al (2001) the fact that gender differences favouring females are restricted to younger adolescents apparently represents a failing of the moral reasoning research to account for gender differences in antisocial behaviour. However, the fact that these gender differences emerge at the time when one starts to see clear evidence of gender divergence in criminal activity may in fact be key to furthering our understanding of the relationship between moral reasoning and offending behaviour. Future research in the field that focuses on gender differences in moral reasoning amongst younger adolescents, and that also pays particular attention to potential sources of variation within samples would help to shed light on this issue.

4.3 CONCLUSION

The empirical research presented in this thesis involved a comprehensive investigation of the patterns of moral reasoning of male offending and male and female non-offending adolescents in the UK using the SRM-SF. By use of differential techniques of analysis a number of theoretical and methodological issues came to light that had not been identified in previous research using the measure.

A major aim of this thesis was to shed light on the controversy surrounding "structural consistency" of moral reasoning with reference to differential techniques of analysis. It was noted that the conventional strategy of treating moral reasoning scores as continuous data and presenting findings with reference to summary mean scores may have served to "mask" heterogeneity in patterns of stage use within and between samples. By adopting alternative "categorical" strategies to tackle SRM-SF data and comparing findings obtained with those obtained using "conventional" techniques, the researcher successfully illustrated that "masking of heterogeneity" was indeed an issue that may have impacted on the interpretation of previous research findings using the measure.

One of the more important messages to be drawn from the current study is that the statistical method used to analyse moral reasoning data clearly <u>does</u> impact on the nature of the findings that emerge. In Chapter 2, when using traditional strategies adopted by researchers to analyse between-group differences in moral reasoning, findings, broadly speaking, corroborated with those presented in previous research studies. However, in Chapter 3, once the data were analysed using alternative techniques several inconsistencies emerged. These were perhaps most notable when comparing between group differences amongst male and female non-offending adolescents. In Chapter 2 when data were analysed with reference to summary means there were no apparent differences between male and female non-offending adolescents, however, in Chapter 3, once the data were reanalysed using alternative techniques clear gender differences emerged. These differences apparently corresponded to the Gilligan (1982) hypothesis that females were making greater use of Stage 3 type reasoning and that males were making greater use of Stage 4 type reasoning. Whilst the findings relating to gender differences may have been idiosyncratic to the current study, the finding that

different techniques of analysis result in different patterns of results is not reconcilable in this way. Researchers planning future work in this area should be aware of the potential impact that differential statistic techniques may have on results that emerge, and should take account of this when interpreting their findings relating to SRM-SF data.

Adopting differential techniques of analysis of SRM-SF data proved useful as a means of gaining insight into potential sources of differences in patterns of moral reasoning between offending adolescents and non-offending adolescents. When comparing overall SRM-SF means of offenders and non-offenders in Chapter 2, offenders demonstrated significantly lower mean scores than non-offending adolescents. This finding was in line with that observed in previous studies. However, on further investigation it was found that characterising the offenders as being morally immature with reference to their non-offending peers was an over-simplistic representation. As a result of subsequent analyses it emerged that the majority of offenders seemed to have the cognitive facility to access moral maturity as evidenced by providing a "Stage 3" justification to at least one SRM-SF item. It was noted that the emphasis on "global cognitive deficits" apparently exhibited by young offenders that seems to be prevalent in the research literature may be counter-productive when applied to clinical interventions that seek to enhance the moral reasoning of young people. In the author's opinion there is a need for future moral reasoning research to pay more attention to when and where young offenders may be performing as well as their non-offending peers. Striving to find ways that areas of specific moral competence can be generalised to other areas would seem to be a more constructive way of approaching clinical interventions with young offenders than focusing on the apparent global inadequacies of the young people concerned.

A novel approach that was adopted in the current study was the attempt to make a systematic investigation of sources of intra-individual variability in patterns of moral stage use with reference to SRM-SF data. It had been noted that previous research studies had, in general, focused on between-group differences in patterns of stage use and that little attention had been paid to potential sources of variation within participant samples. Investigations of patterns of intra-individual variation were

carried out, and it was found that there was evidence of intra-individual variability in patterns of stage use, however, identifying the potential sources of this variability proved problematic.

The development of the SRM-SF as a research tool has made a considerable contribution in terms of potentially furthering our understanding of the moral reasoning of young people. Its user-friendly format makes it practically applicable in a wide range of settings. Its use of open-ended questions provide the researcher with a potentially rich source of data about the way in which people think about moral issues. The SRM-SF was first published in 1992, and, understandably, much of the research carried out over the last decade has focused on establishing its reliability and validity in a variety of settings. When constructing and validating a measure of this nature it is of course essential to establish global benchmarks that can usefully be applied across studies. Indeed, in the current study, one of the primary aims of the first set of analyses presented was to corroborate previous findings using the measure. However, now that there are several published studies indicating that the SRM-SF is globally applicable, in the author's opinion the research focus should now turn to more detailed analysis of the potential sources of individual variation in moral reasoning within the measure.

REFERENCES:

Aleixo, P.A. & Hollin, C.R. (1996) Response bias and lay conceptions of personality and moral reasoning in offenders. *Legal and Criminological Psychology* **1,** 229-232

Arbuthnot, J. & Gordon, D.A. (1986) Behavioral and cognitive effects of a moral reasoning development intervention for high-risk behavior-disordered adolescents. *Journal of Consulting and Clinical Psychology* **34:** 208-216

Barriga, A.Q., Landau, J.R., Stinson, B.L. Liau, A.K. & Gibbs, J.C. (2000) Cognitive distortion and problem behaviors in adolescents. *Criminal Justice and Behavior* **27 (1),** 36-56.

Barriga, A.Q., Morrison, E.M., Liau, A.K. & Gibbs, J.C. (2001) Moral cognition: explaining the gender differences in antisocial behavior. *Merrill-Palmer Quarterly* **47 (4),** 532-562.

Basinger, K.S., Gibbs, J.C. & Fuller, D. (1995) Context and the measurement of moral judgement. *International Journal of Behavioral Development,* **18 (3),** 537-556

Baumrind, D. (1986). Sex differences in moral reasoning: response to Walker's 1984 conclusion that there are none. *Child Development* **57,** 511-21.

Bear, G.G. & Rhys, G.S (1994) Moral reasoning, classroom behavior, and sociometric status among elementary school children. *Developmental Psychology* **30 (5),** 633-638.

Bear, G. G. & Stewart, M. (1990) Early Adolescents' Acceptability of Interventions: Influence of Problem Severity, Gender, and Moral Development. *Journal of Early Adolescence,* **10 (2),** 191-208.

Berkowitz, M.W., Kahn, J. P., Mulry, G. & Piette, (1995) *Psychological and philosophical considerations of prudence and morality.* In M. Killen & D. Hart (eds) Morality in everyday life: Developmental perspectives. Cambridge. CUP.

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

Blatt, M.M. & Kohlberg, L. (1975) The effects of classroom moral discussion upon children's level of moral judgment. *Journal of Moral Education* **4 (2)** 129-161

Blud, L., Travers, R., Nugent, F. & Thornton, D. (2003) Accreditation of offending behaviour programmes in HM Prison Service: "What Works" in practice. *Legal and Criminological Psychology* **8,** 69-81

Bouhmama, D. (1984) Assessment of Kohlberg's Stages of Moral Development in Two Cultures. *Journal of Moral Education* **13 (2)** 124-132

Burton, R.V. (1976) Honesty and Dishonesty: Chapter 10 in T. Lickona (Ed) *Moral Development and Behavior: Theory, Research and Social Issues.* New York. Holt, Rinehart and Winston, 173-197.

Bussey, K. & Maughan, B. (1982) Gender differences in moral reasoning. *Journal of Personality and Social Psychology* 42 (4), 701-706.

Carpendale, J.I.M. & Krebs, D.L. (1992) Situational variation in moral judgement: In a stage or on a stage? *Journal of Youth and Adolescence* **21(2)** 203-224

Carroll, A., Hattie, J., Durkin, K. & Houghton, S. (2001) Goal-setting and reputation enhancement: Behavioural choices among delinquent, at-risk and not at-risk adolescents. *Legal and Criminological Psychology* **6 (2)** 165-184

Chandler, M. & Moran, T. (1990) Psychopathy and moral development: A comparative study of delinquent and nondelinquent youth. *Development and Psychopathology* **2,** 227-246.

Colby, A. & Kohlberg, L. (Eds) (1987) *The Measurement of Moral Judgment. Volume 1: Theoretical Foundations and Research Validation.* New York. Cambridge University Press.

Coleman, J. C. & Hendry, L.B. (1999) *The Nature of Adolescence. Third Edition.* London. Routledge

Coleman, J. C. & Schofield, J (2001) *Key Data on Adolescence.* Brighton. Trust for the Study of Adolescence.

Colledge, M. Collier, P. & Brand, S. (1999) *Programmes for Offenders: Guidance for Evaluators (Crime reduction programme and constructive regimes in prisons – guidance note 2).* Research Development and Statistics Directorate. London. Home Office.

Comunian, A.L. & Gielen, U. P. (1995) Moral Reasoning and Prosocial Action in Italian Culture. *The Journal of Social Psychology* **135 (6)** 699-706

Comunian, A.L. & Gielen, U.P. (2000) Sociomoral reflection and prosocial and antisocial behavior: Two Italian studies. *Psychological Reports* **87** 161-175

Crick, N.R., Casas, J.F. & Ku, H.C. (1999) Relational and physical forms of peer victimization in preschool. *Developmental Psychology* **35 (2),** 376-385.

Elm, D.R., Kennedy, E.J. & Lawton, L. (2001) Determinants of moral reasoning: Sex role orientation, gender and academic factors. *Business and Society* **40 (3)** 241-265.

Emler, N. & Reicher, S. (1995) *Adolescence and Delinquency.* Oxford. Blackwell.

Ferguson, N. & Cairns, E. (1996) Political violence and moral maturity in Northern Ireland. *Political Psychology*, **17 (4),** 713-725.

Ferguson, N., McLernon, F. & Cairns, E (1994) The Sociomoral Reflection Measure – Short Form: an examination of its reliability and validity in a Northern Irish Setting. *British Journal of Educational Psychology*, **64**, 483-489

Ferguson, N., Willis, C.S. & Tilley, A. (2001) Moral Reasoning among Nigerian and Northern Irish children: a cross-cultural comparison using the Sociomoral Reflection Measure – Short Form. *IFE Psychologia* **9,** 1-6

Flood-Page, C., Campbell, S., Harrington, V. & Miller, J. (2000) *Youth crime: Findings from the 1998/1999 Youth Lifestyles Survey* (Home Office Research Study 209). London. Home Office.

Ford, M.R. & Lowery, C.R. (1986) Gender differences in moral reasoning: a comparison of the use of justice and care orientations. Journal of Personality and Social Psychology **50 (4),** 777-783.

Friendship, C., Beech, A.R. & Browne, K.D. (2002a) Reconviction as an outcome measure in research: a methodological note. *British Journal of Criminology* **42,** 412-444.

Friendship, C., Blud, L., Erikson, M. & Travers, R. (2002b) *An evaluation of cognitive behavioural treatment for prisoners.* Home Office Research, Development and Statistics Directorate Findings No. 161. London. Home Office.

Friendship, C., Blud, L. Erikson, M., Travers, R. & Thornton, D. (2003a) Cognitive-behavioural treatment for imprisoned offenders: An evaluation of HM Prison Service's cognitive skills programmes. *Legal and Criminological Psychology* **8 (1),** 103-114

Friendship, C., Falshaw, L. & Beech, A.R. (2003b) Measuring the real impact of accredited offending behaviour programmes. *Legal and Criminological Psychology* **8(1),** 115-127

Galotti, M., Kozberg, S.F. Farmer, M. (1991). Gender and developmental differences in adolescents' conceptions of moral reasoning. *Journal of Youth and Adolescence* **20(1),** 13-30

Garmon, L.C., Basinger, K.S., Gregg, V.R. & Gibbs, J.C. (1996) Gender differences in stage and expression of moral judgment. *Merrill-Palmer Quarterly* **42 (3),** 418-437

Gibbs, J.C. (1977) Kohlberg's stages of moral judgment: a constructive critique. *Harvard Educational Review* **47 (1),** 43-61.

Gibbs, J.C. (1991) Toward an integration of Kohlberg's and Hoffman's moral development theories. *Human Development* **34,** 88-104.

Gibbs, J.C. (2000) Motivating youth to care. *The Ohio, Fall issue.*

Gibbs, J.C., Arnold, K.D., Ahlborn, H.H. & Cheesman, F.L. (1984) Facilitation of sociomoral reasoning in delinquents. *Journal of Consulting and Clinical Psychology* **52** (1) 37-45.

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

Gibbs, J.C., Potter, G.B. & Goldstein, A.P. (1995) *The EQUIP Program: Teaching Youth to Think and Act Responsibly through a Peer-Helping Approach.* Champaign, IL. Research Press.

Gibbs, J.C. & Schnell, S.V. (1985) Moral development "versus" socialization: a critique. *American Psychologist* **40,** 1071-1080

Gibbs, J.C., Widaman, K.F. & Colby, A. (1982) Construction and validation of a simplified, group-administerable equivalent to the Moral Judgment Interview. *Child Development* **53,** 895-910.

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

Gilligan, C. (1998) Remembering Larry. *Journal of Moral Education* **27 (2)** 125-141

Gilligan, C. & Wiggins, G. (1987) The origins of morality in early childhood relationships. In J. Kagan & S. Lamb (ed) *The Emergence of Morality in Young Children.* Chicago. The University of Chicago Press.

Goldstein, A.P., Glick, B. & Gibbs, J.C. (1998) *Aggression Replacement Training – Revised Edition. A Comprehensive Intervention for Aggressive Youth.* Champaign, Illinois. Research 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,** No. 4, 538-553

Haines, K. & Drakeford, M. (1998) *Young People and Youth Justice.* Basingstoke. Palgrave.

Haviv, S. & Leman, P.J. (2002) Moral decision making in real life: Factors affecting moral orientation and behaviour justification. *Journal of Moral Education* **31 (2),** 121-139

Hollin, C.R. (1990) Cognitive-behavioural interventions with young offenders – in custody? *Issues in Criminological Psychology* **15,** 29-38

Hollin, C.R. (1993) Advances in the psychological treatment of delinquent behaviour. *Criminal Behaviour and Mental Health* **3,** 142-157

Hollin, C.R. (1995) The meaning and implications of programme integrity. Chapter 10 in J. McGuire (ed), *What Works? Reducing Reoffending – Guidelines from Research and Practice* (pp 195-208). Chichester. John Wiley & Sons.

Hollin, C.R., Browne, D. & Palmer, E.J. (2002) *Delinquency and Young Offenders (Parent, Adolescent and Child Training Skills 2).* Oxford. BPS Blackwell.

Humphries, M.L., Parker, B. & Jagers, R.J. (2000) Predictors of moral reasoning among African American children: a preliminary study. *Journal of Black Psychology* **26** (1), 51-64.

Izzo, R., & Ross, R.R. (1990) Meta-analysis of rehabilitation programs for juvenile delinquents. *Criminal Justice and Behavior* **17,** No. 1: 134-142

Jadack, R.A., Hyde, J.S., Moore, C.F. & Keller, M.L. (1995) Moral reasoning about sexually transmitted diseases. *Child Development* **66** 167-177.

Jaffee, S. & Hyde, J.S. (2000) Gender differences in moral orientation: A meta-analysis. *Psychological Bulletin* **126 (5)** 703-726

Jamieson, J., McIvor, G. & Murray, C. (1999) *Understanding offending among young people.* Edinburgh. The Stationery Office.

Jurkovic, G.J. (1980) The juvenile delinquent as a moral philosopher: a structural developmental perspective. *Psychological Bulletin* **88 (3)** 709-727.

Kohlberg, L. (1976) *Moral Stages and Moralization: The Cognitive Developmental Approach. In T. Lickona (ed)* Moral Development and Behavior: Theory, Research and Social Issues. New York. Holt, Rinehart & Winston.

Kohlberg, L. (1981) *Essays on Moral Development, Volume 1. The Philosophy of Moral Development: Moral Stages and the Idea of Justice.* San Francisco. Harper & Row.

Kohlberg, L. (1984) *The Psychology of Moral Development: Essays on Moral Development (Vol. 2).* San Francisco. Harper & Row.

Kohlberg, L. & Elfenbein, D. (1975) The development of moral judgements concerning capital punishment. *American Journal of Orthopsychiatry* **45(4)** 614-640

Kohlberg, L., Kauffman, K., Scharf, P. & Hickey, J. (1975) The Just Community approach to corrections: a theory. *Journal of Moral Education* **4 (3)** 243-260

Kohlberg, L., Scharf, P. & Hickey, J. (1971) The justice structure of the prison – a theory and an intervention. *Prison Journal* **51,** 3-14.

Krcmar, M. & Valkenburg, P.M. (1999) A scale to assess children's moral interpretations of justified and unjustified violence and its relationship to television viewing. *Communication Research* **26** (5) 608-634.

Krebs, D.L., Denton, K., Vermeulen, S.C., Carpendale, J.I. & Bush, A. (1991) Structural flexibility of moral judgement. *Journal of Personality and Social Psychology* **61 (6)** 1012-1023.

Krebs, D.L., Denton, K. & Wark, G. (1997). The forms and functions of real-life moral decision making. *Journal of Moral Education* **26 (2)** 131-145.

Krebs, D.L., Vermeulen, S.C., Denton, K. & Carpendale, J.I. (1994). Gender and perspective differences in moral judgement and moral orientation. *Journal of Moral Education* **(23) 1**, 17-26.

Krettenauer, T. & Becker, G. (2001) Entwicklungsniveaus sozio-moralischen Denkens: Deutschspragische Version des Sociomoral Reflection Measure – Short Form (Developmental levels of sociomoral reasoning: German version of the Sociomoral Reflection Measure-Short Form). *Diagnostica* **47 (4)** 188-195

Kuther, T.L. & Higgins-D'Alessandro (2000). Bridging the gap between moral reasoning and adolescent engagement in risky behavior. *Journal of Adolescence* **23,** 409-422.

Lee, M. & Prentice, N.M. (1988) Interrelations of empathy, cognition, and moral reasoning with dimensions of juvenile delinquency. *Journal of Abnormal Child Psychology* **16 (2),** 127-139.

Lipsey, M.W. (1995) What do we learn from 400 research studies on the effectiveness of treatment with juvenile delinquents? Chapter 3 in J. McGuire (ed), *What Works? Reducing Reoffending Guidelines from Research and Practice* (pp 63-78). Chichester. John Wiley & Sons.

Mason, M. & Gibbs, J.C. (1993) Social perspective taking and moral judgement among college students. *Journal of Adolescent Research* **8** (1), 109-123

McGuire, J. (Ed) (1995) *What Works? Reducing Reoffending – Guidelines from Research and Practice*. Chichester. John Wiley & Sons.

McGuire, J. (2000) An introduction to theory and research. Cognitive-behavioural approaches. HM Inspectorate of Probation Report. London. Home Office

McGuire, J. & Priestley, P. (1985) *Offending Behaviour: Skills and Stratagems for Going Straight.* London. B.T. Batsford Ltd.

McGuire, J. & Priestley, P. (1995) Reviewing "what works": past present and future.. Chapter 1 in J. McGuire (ed), *What Works? Reducing Reoffending – Guidelines from Research and Practice* (pp 3-34). Chichester. John Wiley & Sons.

Mizuno, S. (1999) Psychosocial development and moral development: An exploratory comparison of adolescents in Japan and America. *Psychological Reports* **84,** 51-62

Narvaez, D., Getz, I., Rest, J.R. & Thoma, S.J. (1999) Individual moral judgment and cultural ideologies. *Developmental Psychology* **35 (2)** 478-488.

Nelson, J.R., Smith, D.J. & Dodd, J. (1990) The moral reasoning of juvenile delinquents: A meta-analysis. *Journal of Abnormal Child Psychology* **18, (3)** 231-239

Nisan M. & Kohlberg, L. (1982) Universality and variation in moral judgement: a longitudinal and cross-sectional study in Turkey. *Child Development* **53,** 865-876.

Palmer, E.J. & Hollin, C. R. (1997) The influence of perceptions of own parenting on sociomoral reasoning, attributions for criminal behaviour and self-reported delinquency. *Personality and Individual Differences,* **23, (2)** 193-197

Palmer, E.J. & Hollin, C. R. (1998) A comparison of patterns of moral development in young offenders and non-offenders. *Legal and Criminological Psychology,* **3,** 225-235.

Palmer, E. J. & Hollin, C. R. (1999) Social competence and sociomoral reasoning in young offenders. *Applied Cognitive Psychology*, **13,** 79-87.

Palmer, E.J. & Hollin, C.R. (2000) The interrelations of socio-moral reasoning, perceptions of own parenting and attributions of intent with self-reported delinquency. *Legal and Criminological Psychology* **5**, 201-218

Palmer, E.J. & Hollin, C.R. (2001) Sociomoral reasoning, perceptions of parenting and self-reported delinquency in adolescents. *Applied Cognitive Psychology* **15,** 85-100

Putnins, A.L. (1997) Victim Awareness Programs for delinquent youths: Effects on moral reasoning maturity. *Adolescence* **32** (127) 709-714.

Rest, J.R., Narvaez, D., Bebeau, M.J. & Thoma, S.J. (1999) *Postconventional moral thinking: A neo-Kohlbergian approach.* Mahwah, NJ. Lawrence Erlbaum Associates.

Ross, R.R., Fabiano, E.A. & Ewles, C.A. (1988) Reasoning and Rehabilitation. *International Journal of Offender Therapy and Comparative Criminology* **32,** 29-35.

Rutter, M., Giller, H. & Hagell, A. (1998) *Antisocial Behavior by Young People.* Cambridge. Cambridge University Press

Scharf, P. & Hickey, J. (1976) The prison and the inmate's conception of legal justice. *Criminal Justice and Behavior* **3** (2) 107-122

Snarey, J.R. (1985) Cross-cultural universality of social-moral development: A critical review of Kohlbergian research. *Psychological Bulletin* **97 (2)** 202-252.

Snarey, J.R. (1987) Promoting moral maturity among adolescents: An ethnographic study of the Israeli Kibbutz. *Comparative Educational Review* **31 (2)** 241-259

Snarey, J. R., Reimer, J. & Kohlberg, L. (1985a) Development of social-moral reasoning among Kibbutz adolescents: A longitudinal cross-cultural study. *Developmental Psychology,* **21** (1), 3-17

Snarey, J.R., Reimer, J. & Kohlberg, L. (1985b) The Kibbutz as a model for moral education: a cross-cultural study. *Journal of Applied Developmental Psychology* **6**, 151-172

Snyder, S.S. & Feldman, D.H. (1984) Phases of transition in cognitive development: Evidence from the domain of spatial representation. *Child Development* **55**, 981-989

Strauss, A. & Corbin, J. (1998) *Basics of qualitative research: Techniques and procedures for developing grounded theory. Second edition.* Thousand Oaks CA. Sage Publications.

Sykes, G.M. & Matza, D. (1957) Techniques of neutralization: a theory of delinquency. *American Sociological Review,* **22** 664-670

Tangney, J.P. & Dearing, R.L. (2002) Gender differences in morality. Chapter 7 in R.F. Bornstein & J.F. Masling (eds): *The psychodynamics of gender and gender role. Washington DC. American Psychological Association.*

Taylor, J.H. & Walker, L.J. (1997) Moral climate and the development of moral reasoning; the effects of dyadic discussions between young offenders. *Journal of Moral Education* **26** (1), 21-43

Thoma, S.J. & Rest, J.R. (1999) The relationship between moral decision making and patterns of consolidation and transition in moral judgement development. *Developmental Psychology* **35 (2)** 322-344

Van Ijzendoorn, M.H. & Zwart-Woudstra, H.A. (1995) Adolescents' attachment representations and moral reasoning. *Journal of Genetic Psychology* **156** (3) 359-372

Vanstone, M. (2000) Cognitive behavioural work with offenders in the UK: A history of influential endeavour. *The Howard Journal of Criminal Justice* **39 (2),** 171-183.

Vennard, J. Sugg, D. & Hedderman, C. (1997) Part 1, The use of cognitive-behavioural approaches with offenders: messages from the research. *Changing Offenders' Attitudes and Behaviour: What Works?* Home Office Research Study No. 171. London. Home Office

Walker, L.J. (1984). Sex differences in the development of moral reasoning: A critical review. Child Development **55,** 677-91.

Walker, L.J. (1986). Sex differences in the development of moral reasoning: A rejoinder to Baumrind. *Child Development* **57,** 522-526

Walker, L.J. (1989). A longitudinal study of moral reasoning. *Child Development* **60,** 157-166

Walker, L.J., de Vries, B. & Trevethan, S.D. (1987) Moral stages and moral orientations in real-life and hypothetical dilemmas. *Child Development* **58,** 842-858

Walker, L.J., Gustafson, P. & Hennig, K.H. (2001) The consolidation/transition model in moral reasoning development. *Developmental Psychology* **37 (2),** 187-197

Walker, L.J. & Moran, T.J. (1991) Moral reasoning in a Communist Chinese society. *Journal of Moral Education* **20 (2),** 139-155

Walker, L.J. & Taylor, J.H. (1991a) Family interactions and the development of moral reasoning. *Child Development* **62,** 264-283

Walker, L.J. & Taylor, J.H. (1991b) Stage transitions in moral reasoning: A longitudinal study of developmental processes. *Developmental Psychology* **27 (2),** 330-337.

Wark, G.R. & Krebs, D.L. (1996) Gender and dilemma differences in real-life moral judgment. *Developmental Psychology* **32 (2)** 220-230

Wark, G.R. & Krebs, D.L. (2000) The construction of moral dilemmas in everyday life. *Journal of Moral Education* **29 (1)** 5-21

Watt, M.C., Frausin, S. Dixon, J & Nimmo, S. (2000) Moral intelligence in a sample of incarcerated females. *Criminal Justice and Behavior* **27 (3),** 330-355

Wilson, S., Attrill, G. & Nugent, F. (2003) Effective interventions for acquisitive offenders: An investigation of cognitive skills programmes. *Legal and Criminological Psychology* **8 (1),** 83-101

APPENDICES

Appendix 1: The Sociomoral Reflection Measure-Short Form.

Instructions

Date of Birth:

Gender:

Instructions

In this questionnaire we want to find out about the things that you think are important for people to do, and especially why you think these things are important. Please try to help us understand your thinking by WRITING AS MUCH AS YOU CAN TO EXPLAIN. Please don't just write "same as before. Please answer ALL the questions, especially the "Why" questions. Thank you very much for your help.

1. Think about when you have made a promise to a friend of yours. How important is it for people to keep promises, if they can, to friends?

Circle one: very important important not important

WHY IS THAT VERY IMPORTANT / IMPORTANT / NOT IMPORTANT

(WHICHEVER ONE YOU CIRCLED)?

2. How about keeping a promise to anyone? How important is it for people to keep promises, if they can, even to someone they hardly know?

Circle one: very important important not important

WHY IS THAT VERY IMPORTANT / IMPORTANT / NOT IMPORTANT

(WHICHEVER ONE YOU CIRCLED)?

3. How about keeping a promise to a child? How important is it for parents to keep promises, if they can to their children?

Circle one: very important important not important

WHY IS THAT VERY IMPORTANT / IMPORTANT / NOT IMPORTANT

(WHICHEVER ONE YOU CIRCLED)?

4. In general, how important is it to tell the truth?

Circle one: very important important not important

WHY IS THAT VERY IMPORTANT / IMPORTANT / NOT IMPORTANT

(WHICHEVER ONE YOU CIRCLED)?

5. Think about when you've helped your mother or father. How important is it for children to help their parents?

Circle one: very important important not important

WHY IS THAT VERY IMPORTANT / IMPORTANT / NOT IMPORTANT

(WHICHEVER ONE YOU CIRCLED)?

6. Let's say a friend of yours needs help and may even die, and you're the only person who can save him or her. How important is it for a person (without losing his or her own life) to save the life of a friend?

Circle one: very important important not important

WHY IS THAT VERY IMPORTANT / IMPORTANT / NOT IMPORTANT

(WHICHEVER ONE YOU CIRCLED)?

7. What about saving the life of anyone? How important is it for a person (without losing his or her own life) to save the life of a stranger?

Circle one: very important important not important

WHY IS THAT VERY IMPORTANT / IMPORTANT / NOT IMPORTANT

(WHICHEVER ONE YOU CIRCLED)?

8. How important is it for a person to live even if that person doesn't want to?

Circle one: very important important not important

WHY IS THAT VERY IMPORTANT / IMPORTANT / NOT IMPORTANT

(WHICHEVER ONE YOU CIRCLED)?

9. How important is it for people not to take things that belong to other people?

Circle one: very important important not important

WHY IS THAT VERY IMPORTANT / IMPORTANT / NOT IMPORTANT

(WHICHEVER ONE YOU CIRCLED)?

10. How important is it for people to obey the law?

Circle one: very important important not important

WHY IS THAT VERY IMPORTANT / IMPORTANT / NOT IMPORTANT

(WHICHEVER ONE YOU CIRCLED)?

11. How important is it for judges to send people who break the law to jail?

Circle one: very important important not important

WHY IS THAT VERY IMPORTANT / IMPORTANT / NOT IMPORTANT

(WHICHEVER ONE YOU CIRCLED)?

Appendix 2. Pearson Correlation Coefficients Obtained from Comparisons of SRM-SF Items and SRM-SF Sociomoral Values

2a) Pearson correlation coefficients for inter-item comparisons

		Q1 Promise Friend	Q2 Promise Stranger	Q3 Promise Children	Q4 Telling Truth	Q5 Helping Parents	Q6 Save Friend	Q7 Save Stranger	Q8 Stay Alive	Q9 Not Steal	Q10 Obey Law	Q11 Judges Jailing
Q1												
Q2	r	.501										
	p	.000										
	N	294	311									
Q3	r	.353	.442									
	p	.000	.000									
	N	325	267	353								
Q4	r	.245	.324	.385								
	p	.000	.000	.000								
	N	330	268	305	356							
Q5	r	.312	.347	.326	.291							
	p	.000	.000	.000	.000							
	N	331	267	307	313	362						
Q6	r	.335	.387	.364	.309	.331						
	p	.000	.000	.000	.000	.000						
	N	309	257	288	290	296	340					
Q7	r	.210	.234	.227	.226	.213	.264					
	p	.001	.000	.000	.000	.001	.000					
	N	236	202	221	222	221	222	256				
Q8	r	.255	.236	.256	.327	.340	.303	.237				
	p	.000	.000	.000	.000	.000	.000	.000				
	N	273	225	257	260	263	243	192	301			
Q9	r	.322	.364	.216	.132	.267	.258	.200	.207			
	p	.000	.000	.000	.011	.000	.000	.002	.001			
	N	323	263	296	299	306	288	217	247	349		
Q10	r	.335	.296	.380	.484	.384	.288	.270	.311	.256		
	p	.000	.000	.000	.000	.000	.000	.000	.000	.000		
	N	293	241	275	278	280	270	200	237	270	322	
Q11	r	.340	.360	.342	.338	.350	.233	.277	.378	.296	.529	
	p	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
	N	279	227	256	265	264	253	184	221	259	241	

2b) Pearson correlation coefficients obtained from inter-value comparisons

		CONTRUTH	AFFIL	LIFE	PROP_LAW	JUSTICE
CONTRUTH	Pearson Correlation					
	Sig. (1-tailed)					
	N					
AFFIL	Pearson Correlation	.494				
	Sig. (1-tailed)	.000				
	N	405				
LIFE	Pearson Correlation	.331	.358			
	Sig. (1-tailed)	.000	.000			
	N	364	359			
PROP_LAW	Pearson Correlation	.474	.360	.288		
	Sig. (1-tailed)	.000	.000	.000		
	N	400	394	353		
JUSTICE	Pearson Correlation	.443	.337	.361	.458	
	Sig. (1-tailed)	.000	.000	.000	.000	
	N	304	299	264	298	

Appendix 3 Overall Frequencies of Moral Stage Use by Participant Group:

	Number of Responses															
Stage 1	0		1		2		3		4		5		6		7+	
	(n)	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)	%
offender	75	51.0	44	29.9	19	12.9	5	3.4	3	2.0	1	0.7				
male non-offender	98	65.8	38	25.5	7	4.7	3	2.0	3	2.0						
female non-offender	100	85.5	15	12.8	2	1.7										
Total	273	66.1	97	23.5	28	6.8	8	1.9	6	1.5	1	0.2				
Transition Stage 1-2	0		1		2		3		4		5		6		7+	
	n	%	n	%	n	%	n	%	n	%	n	%				
Offenders	60	40.8	57	38.8	21	14.3	6	4.1	3	2.0						
Male Non-Offenders	92	61.7	44	29.5	11	7.4	1	0.7	1	0.7						
Female Non-Offenders	77	65.8	31	26.5	7	6.0	2	1.7								
All	229	55.4	132	32.0	39	9.4	9	2.2	4	1.0						
Stage 2																
Frequency	0		1		2		3		4		5		6		7+	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Offenders	9	6.1	26	17.7	34	23.1	36	24.5	21	14.3	11	7.5	7	4.8	3	2.1
Male Non-Offenders	45	30.2	38	25.5	31	20.8	23	15.4	8	5.4	2	1.3	1	0.7	1	0.7
Female Non-Offenders	25	21.4	39	33.3	26	22.2	16	13.7	6	5.1	3	2.6	2	1.7		
All	79	19.1	103	24.9	91	22.0	75	18.2	35	8.5	16	3.9	10	2.4	4	0.9
Stage 2-3																
Frequency	0		1		2		3		4		5		6		7+	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Offenders	10	6.8	22	15.0	37	25.2	35	23.8	24	16.3	13	8.8	5	3.4	1	0.7
Male Non-Offenders	26	17.4	36	24.2	36	24.2	26	17.4	18	12.1	5	3.4	2	1.3		
Female Non-Offenders	4	3.4	24	20.5	34	29.1	22	18.8	19	16.2	10	8.5	3	2.6	1	0.9
All	40	9.7	82	19.9	107	25.9	83	20.1	61	14.8	28	6.8	10	2.4	2	0.5
Stage 3	0		1		2		3		4		5		6		7+	
Frequency	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Offenders	38	25.9	45	30.6	31	21.1	23	15.6	6	4.1	2	1.4	2	1.4		
Male Non-Offenders	26	17.4	30	20.1	28	18.8	35	23.5	14	9.4	13	8.7	3	2.0		
Female Non-Offenders	9	7.7	14	12.0	38	32.5	31	26.5	9	7.7	9	7.7	7	6.0		
All	73	17.7	89	21.5	97	23.5	89	21.5	29	7.0	24	5.8	12	2.9		
Stage 3-4	0		1		2		3		4		5		6		7+	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Offenders	127	86.4	12	8.2	6	4.1			1	0.7			1	0.7		
Male Non-Offenders	47	31.5	37	24.8	33	22.1	14	9.4	8	5.4	6	4.0	3	2.0	1	0.7
Female Non-Offenders	49	41.9	33	28.2	16	13.7	9	7.7	5	4.3	2	1.7	1	0.9	2	1.8
All	223	54.0	82	19.9	55	13.3	23	5.6	14	3.4	8	1.9	5	1.2	3	0.7
Stage 4	0		1		2		3		4		5		6		7+	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Offenders	143	97.3	3	2.0			1	0.7								
Male Non-Offenders	107	71.8	27	18.1	6	4.0	5	3.4	3	2.0	1	0.7				
Female Non-Offenders	109	93.2	8	6.8												
All	359	86.9	38	9.2	6	1.5	6	1.5	3	0.7	1	0.2				

Appendix 4 Frequency Data Used for Chi-Square Analysis of Between-Group Differences in Patterns of Stage Use

	1	1	1-2	1-2	2	2	2-3	2-3	3	3	3-4	3-4	4	4
	(n)	%	n	%	n	%	n	%	n	%	n	%	n	%
Promise Friend O	15	11.5	5	3.8	23	17.7	51	39.2	28	21.5	8	6.2	0	0.0
MN	5	3.6	2	1.5	10	7.2	44	31.9	43	31.2	31	22.5	2	1.4
FN	1	0.9	1	.9	6	5.3	32	28.3	49	43.4	23	20.4	1	0.9
Total	21	5.5	8	2.1	39	10.2	127	33.3	120	31.5	62	16.3	3	0.8
Promise Stranger O	7	6.2	3	2.7	51	45.1	35	31.0	13	11.5	3	2.7	1	0.9
MN	5	4.7	1	1.0	24	22.9	31	29.5	19	18.1	18	17.1	7	6.7
FN	1	1.1	1	1.1	17	18.3	36	38.7	22	23.7	15	16.1	1	1.1
Total	13	4.2	5	1.6	92	29.6	102	32.8	54	17.4	36	11.6	9	2.9
Promise Children O	17	13.7	4	3.2	11	8.9	53	42.7	36	29.0	2	1.6	1	0.8
MN	3	2.3	5	3.8	6	4.6	31	23.7	22	16.8	54	41.2	10	7.6
FN	2	2.0	3	3.1	9	9.2	19	19.4	24	24.5	41	41.8	0	0.0
Total	22	6.2	12	3.4	26	7.4	103	29.2	82	23.2	97	27.5	11	3.1
Telling Truth O	7	5.3	5	3.8	63	47.4	49	36.8	8	6.0	0	0.0	1	0.8
MN	3	2.4	2	1.6	30	24.4	38	30.9	22	17.9	22	17.9	6	4.9
FN	3	3.0	4	4.0	41	41.4	26	26.3	17	17.2	8	8.1	0	0.0
Total	13	3.7	11	3.1	134	37.7	113	31.8	47	13.2	30	8.5	7	2.0
Helping Parents O	8	6.2	2	1.6	45	34.9	48	37.2	19	14.7	7	5.4	0	0.0
MN	3	2.3	1	.8	26	20.0	44	33.8	29	22.3	24	18.5	3	2.3
FN					11	10.7	54	52.4	25	24.3	12	11.7	1	1.0
Total	11	3.0	3	.8	82	22.7	146	40.3	73	20.2	43	11.9	4	1.1
Saving Friend O	17	13.4	3	2.4	23	18.1	40	31.5	43	33.9	1	0.8	0	0.0
MN	14	11.6	1	.8	18	14.9	18	14.9	57	47.1	7	5.8	6	5.0
FN	1	1.1	1	1.1	5	5.4	24	26.1	58	63.0	3	3.3	0	0.0
Total	32	9.4	5	1.5	46	13.5	82	24.1	158	46.5	11	3.2	6	1.8
Saving Stranger O	5	5.3	4	4.2	30	31.6	26	27.4	28	29.5	2	2.1	0	0.0
MN	4	4.2	3	3.1	21	21.9	11	11.5	49	51.0	5	5.2	3	3.1
FN	2	3.1	0	0.0	22	33.8	8	12.3	30	46.2	3	4.6	0	0.0
Total	11	4.3	7	2.7	73	28.5	45	17.6	107	41.8	10	3.9	3	1.2
Staying Alive O	2	1.8	1	0.9	50	44.2	25	22.1	32	28.3	3	2.7	0	0.0
MN	3	3.1	0	0.0	30	30.6	20	20.4	24	24.5	18	18.4	3	3.1
FN	0	0.0	0	0.0	20	22.2	23	25.6	34	37.8	11	12.2	2	2.2
Total	5	1.7	1	0.3	100	33.2	68	22.6	90	29.9	32	10.6	5	1.7
Not Stealing O	13	10.3	41	32.5	30	23.8	34	27.0	6	4.8	2	1.6	0	0.0
MN	23	18.4	22	17.6	17	13.6	27	21.6	28	22.4	5	4.0	3	2.4
FN	6	6.1	17	17.3	11	11.2	43	43.9	15	15.3	6	6.1	0	0.0
Total	42	12.0	80	22.9	58	16.6	104	29.8	49	14.0	13	3.7	3	0.9
Obeying the Law O	14	12.1	31	26.7	41	35.3	20	17.2	5	4.3	3	2.6	2	1.7
MN	5	4.3	13	11.2	18	15.5	9	7.8	28	24.1	23	19.8	20	17.2
FN	2	2.2	12	13.3	14	15.6	22	24.4	25	27.8	12	13.3	3	3.3
Total	21	6.5	56	17.4	73	22.7	51	15.8	58	18.0	38	11.8	25	7.8
Legal Justice O	9	8.7	30	28.8	39	37.5	18	17.3	4	3.8	3	2.9	1	1.0
MN	5	4.3	23	20.0	23	20.0	22	19.1	9	7.8	25	21.7	8	7.0
FN	1	1.2	12	14.0	35	40.7	22	25.6	7	8.1	9	10.5	0	0.0
Total	15	4.9	65	21.3	97	31.8	62	20.3	20	6.6	37	12.1	9	3.0

Appendix 5: Chi Square Statistics for Comparisons of Between-Group Differences in Stage Use Across SRM-SF items.

Stage 1	Offenders / Male Non-Offenders	Male / Female Non-Offenders
Promise Friend	$\chi^2 = 5.99$; df =1; p<0.05	$\chi^2 = 2.02$; df =1; ns[a]
Promise Stranger	$\chi^2 = 0.23$; df =1; ns	$\chi^2 = 2.25$; df =1; ns[a]
Promise Child	$\chi^2 = 11.49$; df =1; p<0.01	$\chi^2 = 0.02$; df =1; ns[a]
Tell Truth	$\chi^2 = 1.36$; df =1; ns[a]	$\chi^2 = 0.07$; df =1; ns[a]
Help Parents	$\chi^2 = 2.41$; df =1; ns	$\chi^2 = 2.41$; df =1; ns[a]
Save Friend	$\chi^2 = 0.19$; df =1; ns	$\chi^2 = 8.77$; df =1; p<0.01
Save Stranger	$\chi^2 = 0.13$; df =1; ns	$\chi^2 = 0.13$; df =1; ns[a]
Stay Alive	$\chi^2 = 0.38$; df =1; ns	$\chi^2 = 2.80$; df =1; ns[a]
Not Steal	$\chi^2 = 3.34$; df =1; ns	$\chi^2 = 7.32$; df =1; p<0.01
Obey Law	$\chi^2 = 4.64$; df =1; p<0.05	$\chi^2 = 0.67$; df =1; ns[a]
Legal Justice	$\chi^2 = 1.69$; df =1; ns	$\chi^2 = 1.72$; df =1; ns[a]

[a]Fisher's Exact Test

Stage 1-2	Offenders / Male Non-Offenders	Male / Female Non-Offenders
Promise Friend	$\chi^2 = 1.49$; df =1; ns[1]	$\chi^2 = 0.17$; df =1; ns[1]
Promise Stranger	$\chi^2 = 0.88$; df =1; ns[1]	$\chi^2 = 0.01$; df =1; ns[1]
Promise Child	$\chi^2 = 0.07$; df =1; ns[1]	$\chi^2 = 0.10$; df =1; ns[1]
Tell Truth	$\chi^2 = 1.09$; df =1; ns[1]	$\chi^2 = 1.22$; df =1; ns[1]
Help Parents	$\chi^2 = 0.35$; df =1; ns[1]	$\chi^2 = 0.80$; df =1; ns[1]
Save Friend	$\chi^2 = 0.92$; df =1; ns[1]	$\chi^2 = 0.04$; df =1; ns[1]
Save Stranger	$\chi^2 = 0.16$; df =1; ns[1]	$\chi^2 = 2.07$; df =1; ns[1]
Stay Alive	$\chi^2 = 0.87$; df =1; ns[1]	[2]
Not Steal	$\chi^2 = 7.45$; df =1; p<0.01	$\chi^2 = 0.00$; df =1; ns
Obey Law	$\chi^2 = 9.09$; df =1; p<0.01	$\chi^2 = 0.22$; df =1; ns
Judges Jail	$\chi^2 = 2.33$ df =1; ns	$\chi^2 = 1.25$; df =1; ns

[1]Fisher's Exact Test
[2]Not computed – none of the non-offenders used this stage for this item

Stage 2	Offenders / Male Non-Offenders	Male / Female Non-Offenders
Promise Friend	$\chi^2 = 6.76$; df =1; p<0.01	$\chi^2 = 0.39$; df =1; ns
Promise Stranger	$\chi^2 = 11.97$; df =1; p<0.01	$\chi^2 = 0.63$; df =1; ns
Promise Child	$\chi^2 = 1.86$; df =1; ns	$\chi^2 = 1.94$; df =1; ns
Tell Truth	$\chi^2 = 14.59$; df =1; p<0.001	$\chi^2 = 7.31$; df =1; p<0.01
Help Parents	$\chi^2 = 7.21$; df =1; p<0.01	$\chi^2 = 3.74$; df =1; ns
Save Friend	$\chi^2 = 0.47$; df =1; ns	$\chi^2 = 4.84$; df =1; p<0.05
Save Stranger	$\chi^2 = 2.30$; df =1; ns	$\chi^2 = 2.84$; df =1; ns
Stay Alive	$\chi^2 = 4.15$; df =1; p<0.05	$\chi^2 = 1.69$; df =1; ns
Not Steal	$\chi^2 = 4.30$; df =1; p<0.05	$\chi^2 = 0.28$; df =1; ns
Obey Law	$\chi^2 = 12.02$; df =1; p<0.01	$\chi^2 = 0.00$; df =1; ns
Judges Jail	$\chi^2 = 8.24$; df =1; p<0.01	$\chi^2 = 10.27$; df =1; p<0.01

Stage 2 - 3	Offenders / Male Non-Offenders	Male / Female Non-Offenders
Promise Friend	$\chi^2 = 1.58$; df =1; ns	$\chi^2 = 0.37$; df =1; ns
Promise Stranger	$\chi^2 = 0.54$; df =1; ns	$\chi^2 = 1.86$; df =1; ns
Promise Child	$\chi^2 = 10.50$; df =1; p<0.01	$\chi^2 = 0.60$; df =1; ns
Tell Truth	$\chi^2 = 1.01$; df =1; ns	$\chi^2 = 0.57$; df =1; ns
Help Parents	$\chi^2 = 0.32$; df =1; ns	$\chi^2 = 8.14$; df =1; p<0.01
Save Friend	$\chi^2 = 9.55$; df =1; p<0.01	$\chi^2 = 4.15$; df =1; p<0.05
Save Stranger	$\chi^2 = 7.74$; df =1; p<0.01	$\chi^2 = 0.03$; df =1; ns
Stay Alive	$\chi^2 = 0.92$; df =1; ns	$\chi^2 = 0.71$; df =1; ns
Not Steal	$\chi^2 = 0.99$; df =1; ns	$\chi^2 = 12.66$; df =1; p<0.001
Obey Law	$\chi^2 = 4.77$; df =1; p<0.05	$\chi^2 = 11.04$; df =1; p<0.01
Judges Jail	$\chi^2 = 0.12$; df =1; ns	$\chi^2 = 1.20$; df =1; ns

Appendix 5: Chi Square Statistics for Comparisons of Between-Group Differences in Stage Use Across SRM-SF items.

Stage 3	Offenders / Male Non-Offenders	Male / Female Non-Offenders
Promise Friend	χ^2 = 3.18; df =1; ns	χ^2 = 3.99; df =1; p<0.05
Promise Stranger	χ^2 = 1.89; df =1; ns	χ^2 = 0.93; df =1; ns
Promise Child	χ^2 = 5.43; df =1; p<0.05	χ^2 = 2.07; df =1; ns
Tell Truth	χ^2 = 8.71; df =1; p<0.01	χ^2 = 0.02; df =1; ns
Help Parents	χ^2 = 2.46; df =1; ns	χ^2 = 0.13; df =1; ns
Save Friend	χ^2 = 4.52; df =1; p<0.05	χ^2 = 5.34; df =1; p<0.05
Save Stranger	χ^2 = 9.23; df =1; p<0.01	χ^2 = 0.37; df =1; ns
Stay Alive	χ^2 = 0.40; df =1; ns	χ^2 = 3.88; df =1 p<0.05
Not Steal	χ^2 = 16.67; df =1; p<0.001	χ^2 = 1.78; df =1; ns
Obey Law	χ^2 = 18.69; df =1; p<0.001	χ^2 = 0.35; df =1; ns
Judges Jail	χ^2 = 1.55; df =1; ns	χ^2 = 0.01; df =1; ns

Stage 3-4	Offenders / Male Non-Offenders	Male / Female Non-Offenders
Promise Friend	χ^2 = 14.32; df =1; p<0.001	χ^2 = 0.16; df =1; ns
Promise Stranger	χ^2 = 13.12; df =1; p<0.001	χ^2 = 0.04; df =1; ns
Promise Child	χ^2 = 58.31; df =1; p<0.001	χ^2 = 0.01; df =1; ns
Tell Truth	χ^2 = 26.03; df =1; p<0.001	χ^2 = 4.51; df =1; p<0.05
Help Parents	χ^2 = 10.44; df =1; p<0.01	χ^2 = 2.04; df =1; ns
Save Friend	χ^2 = 4.96; df =1; p<0.05	χ^2 = 0.74; df =1; ns[1]
Save Stranger	χ^2 = 1.30; df =1; ns[1]	χ^2 = 0.03; df =1; ns[1]
Stay Alive	χ^2 = 14.46; df =1; p<0.001	χ^2 = 1.36; df =1; ns
Not Steal	χ^2 = 1.35; df =1; ns[1]	χ^2 = 0.53; df =1; ns[1]
Obey Law	χ^2 = 17.33; df =1; p<0.001	χ^2 = 1.52; df =1; ns
Judges Jail	χ^2 = 17.41; df =1; p<0.001	χ^2 = 4.45; df =1; p<0.05

Stage 4	Offenders / Male Non-Offenders	Male / Female Non-Offenders
Promise Friend	χ^2 = 1.90; df =1; ns[1]	χ^2 = 0.17; df =1; ns[1]
Promise Stranger	χ^2 = 5.15; df =1; p<0.05[1]	χ^2 = 3.98; df =1; ns[1]
Promise Child	χ^2 = 7.19; df =1; p<0.01	χ^2 = 7.82; df =1; p<0.01[1]
Tell Truth	χ^2 = 4.09; df =1; p ns[1]	χ^2 = 4.96; df =1; p<0.05[1]
Help Parents	χ^2 = 3.02; df =1; ns[1]	χ^2 = 0.61; df =1; ns[1]
Save Friend	χ^2 = 6.44; df =1; p<0.05[1]	χ^2 = 4.69; df =1; p<0.05[1]
Save Stranger	χ^2 = 3.02; df =1; ns[1]	χ^2 = 2.07; df =1; ns[1]
Stay Alive	χ^2 = 3.51; df =1; ns[1]	χ^2 = 0.13; df =1; ns[1]
Not Steal	χ^2 = 3.06; df =1; ns[1]	χ^2 = 2.38; df =1; ns[1]
Obey Law	χ^2 = 16.27; df =1; p<0.001	χ^2 = 9.88; df =1; p<0.01
Judges Jail	χ^2 = 4.98; df =1; p<0.05[1]	χ^2 = 6.23; df =1; p<0.05[1]

[1] *Fisher's Exact Test*