Development and Validation of a Measure to Assess
Moral Reasoning in Singapore Secondary School Students

Lyndon Lim
B.Eng (Hons), PGDip.Ed. (Dist), M.Ed.

Supervisors: Dr. Elaine Chapman and Professor David Andrich

This thesis is presented in partial fulfilment of the requirements for the degree of Doctor of Education of The University of Western Australia

Graduate School of Education, 2017
Thesis Declaration

I, Lyndon Lim, certify that:

(i) This thesis has been substantially accomplished during enrolment in the degree.

(ii) This thesis does not contain material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution.

(iii) No part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of The University of Western Australia and where applicable, any partner institution responsible for the joint-award of this degree.

(iv) This thesis does not contain any material previously published or written by another person, except where due reference has been made in the text.

(v) The work(s) are not in any way a violation or infringement of any copyright, trademark, patent, or other rights whatsoever of any person.

(vi) The research involving human data reported in this thesis was assessed and approved by The University of Western Australia Human Research Ethics Committee. Approval #: RA/4/1/7813.

(vii) Approval to collect data from schools was obtained from the Singapore Ministry of Education prior to commencing the relevant work described in this thesis.

(viii) This thesis does not contain work that I have published, nor work under review for publication.

Lyndon Lim

20 October 2017
Abstract

Moral reasoning is a key element of the Singapore Ministry of Education’s current Character and Citizenship Education (CCE) curriculum. The moral reasoning element of the CCE is structured on the basis of Kohlberg’s (1984) Theory of Moral Development. Teachers in Singapore secondary schools are currently being encouraged to apply this theory in their efforts to facilitate growth in students’ moral reasoning. To monitor students’ progress in this area, however, a reliable and valid test of moral reasoning is needed. While two widely used instruments are already available to assess moral reasoning development based on Kohlberg’s model, neither of these instruments provide a practical means by which large numbers of students can be assessed. Thus, the goal of this study was to develop and provide a preliminary validation of an instrument to assess moral reasoning development in Singapore secondary classrooms, based on the CCE curriculum.

The instrument developed and validated in the present study was named the Moral Reasoning Questionnaire, or MRQ. The MRQ is based on Kohlberg’s (1984) theoretical model, with items presenting moral dilemmas that integrate prosocial, anti-social and social pressure elements within situations that should be familiar to most Singapore secondary level students. The MRQ was developed and validated in a series of stages, as suggested by Messick (1993, 1995). In Stage 1, the MRQ was developed in consultation with the thesis supervisor. Content appropriateness was then reviewed and affirmed by an expert panel. In Stage 2, individual cognitive interviews were conducted with a group of five students, who ‘thought aloud’ as they completed the MRQ. The latter interviews provided evidence to support the response processes associated with the MRQ. Items were
then refined for use in Stage 3, and data \((n=670)\) were collected from three secondary schools. Exploratory and confirmatory factor analyses provided evidence to support the reliability and validity of the MRQ in terms of internal structure. A Rasch analysis also indicated that the MRQ was unidimensional and had acceptable psychometric properties. Based on the latter analysis, however, the initial six categories of scores were collapsed into a smaller subset of three. Multi-group confirmatory factor analyses further supported the assumption of measurement invariance for gender and upper/lower secondary levels for the MRQ, which suggests that this instrument can be used to draw meaningful comparisons between males and females or upper and lower secondary students.

This research was designed to contribute to the professional practices of teachers in Singapore, by providing a standardised and practical measure to assess moral reasoning development in large groups of secondary students. It is suggested that further research be undertaken to determine if the MRQ is measurement invariant when other factors such as culture and social economic status are considered. As the scenarios used in the items should be familiar to most students at the secondary level or equivalent, there is potential for the MRQ to be edited for use in countries with education systems comparable to Singapore’s. An alternative instrument suitable for use in the Australian context, the MRQ-A, has already been prepared by the author to determine measurement invariance due to culture. Further research is currently underway to trial the MRQ-A within a large sample of Australian secondary students.
# Table of Contents

THESIS DECLARATION .......................................................... II

ABSTRACT ........................................................................ III

TABLE OF CONTENTS .......................................................... V

LIST OF FIGURES................................................................ IX

LIST OF TABLES ...................................................................... XI

ACKNOWLEDGEMENTS ........................................................... XIII

CHAPTER 1. THE CONSTRUCT OF MORAL REASONING .............. 1

1.1. ISSUES IN DEFINING MORAL REASONING ......................... 1

1.1.1. Moral reasoning vs. ethics .................................................. 3

1.1.2. Moral reasoning vs. moral intensity and moral orientation ....... 5

1.1.3. Definition adopted within the present study ......................... 8

1.2. THEORETICAL FRAMEWORKS ON MORAL REASONING ........ 8

1.2.1. Piaget's theoretical framework on moral understanding ......... 9

1.2.2. Kohlberg's Stage Theory of Moral Reasoning .................... 13

1.2.3. Frameworks by Rest and colleagues ................................. 19

1.2.4. Theoretical framework by Gilligan ................................. 24

1.2.5. Alternative frameworks and strategies ............................. 28

1.3. CULTURE AND MORAL REASONING .............................. 32

1.4. GENDER AND MORAL REASONING ............................... 37

1.5. THE PRESENT STUDY CONTEXT ........................................ 38

1.6. OVERVIEW OF THE PRESENT STUDY .............................. 40
CHAPTER 2. ASSESSING MORAL REASONING.................................................42

2.1. MORAL JUDGMENT INTERVIEW..........................................................43

2.2. DEFINING ISSUES TEST........................................................................48

2.3 ETHICAL REASONING INVENTORY.........................................................60

2.4. SOCIOMORAL MEASURES .................................................................62

2.5. THE NEED FOR A NEW MORAL REASONING INSTRUMENT...............80

CHAPTER 3. METHOD ....................................................................................87

3.1. MESSICK’S UNITARY CONCEPT OF VALIDITY ..................................87

3.2. PARTICIPANTS.......................................................................................89

3.3. INSTRUMENT DESIGN..........................................................................92

3.3.1. Item stems .......................................................................................92

3.3.2. Response options ..........................................................................94

3.3.3. Item scoring ....................................................................................98

3.4. STAGE 1 DATA COLLECTION PROCEDURES: CONTENT APPROPRIATENESS OF ITEMS - EXPERT JUDGMENT .........................................................98

3.5. STAGE 2 DATA COLLECTION PROCEDURES: EVIDENCE BASED ON RESPONSE PROCESSES ..............................................................................100

3.6. STAGE 3 DATA COLLECTION PROCEDURES: EVIDENCE BASED ON THE INTERNAL STRUCTURE OF THE TEST ..............................................103

CHAPTER 4. RESULTS....................................................................................104

4.1. EXPLORATORY FACTOR ANALYSIS .....................................................105

4.2. CONFIRMATORY FACTOR ANALYSIS .................................................113

4.3. RASCH ANALYSIS OF ITEM SCORES ..................................................120

4.3.1. Threshold Ordering and Reliability ..................................................122
List of Figures

Figure 1. Mapping of moral reasoning to its two functions* .................................. 21
Figure 2. Example of a DIT item* ........................................................................ 50
Figure 3. Example of a DIT2 item* ...................................................................... 55
Figure 4. Example item (1) of the SRM-SFO* ....................................................... 74
Figure 5. Example item (11) of SRM-SFO* .......................................................... 75
Figure 6. Measurement model for non-delinquent sample ...................................... 78
Figure 7. Measurement model for delinquent sample ............................................. 79
Figure 8. Example item in the TMV ...................................................................... 85
Figure 9. Example item with vignette and corresponding options ........................... 97
Figure 10. Document for judges to score and comment ........................................ 99
Figure 11. Cognitive interview model by Ryan et al. (2012) .................................... 102
Figure 12. Scree plot and variance explained figures by ML estimation ............... 108
Figure 13. Scree plot and variance explained figures by ULS estimation ............. 108
Figure 14. Parallel analysis graphical output .......................................................... 109
Figure 15. Measurement model with MLSB estimation ........................................... 119
Figure 16. Threshold map for 26 items ................................................................. 124
Figure 17. ICC of A9 ............................................................................................. 129
Figure 18. ICC of A26 .......................................................................................... 129
Figure 19. ICC with the level plots for A10 (item 9 in RUMM2030) ..................... 132
Figure 20. PCA t-test of +/- loaded items on first principal component ............... 133
Figure 21. Person-item threshold distribution ....................................................... 134
Figure 22. Person-item measure threshold distribution by gender ....................... 135
Figure 23. Person-item measure threshold distribution by school ....................... 135
Figure 24. Person-item measure threshold distribution by stream ..................... 136
Figure 25. Person-item measure threshold distribution by level of study .......... 136
Figure 26. Scree plot and variance explained figures by ML estimation* ........ 138
Figure 27. Parallel analysis for MRQ with collapsed categories ...................... 139
Figure 28. Path diagram with standardised loading estimates ....................... 141
Figure 29. How CAT works* .................................................................................. 158
List of Tables

Table 1. Mapping of Ethical Reasoning to Kohlberg’s Stages*................................. 5
Table 2. Piaget’s Two-phase theory*................................................................. 10
Table 3. Summary of Kohlberg’s Six Stages of Moral Development* ............... 15
Table 4. Example statements against stages of moral development .................. 22
Table 5. Mapping of Kohlberg’s Theory and the Schema Theory ...................... 24
Table 6. Kohlberg’s and Gilligan’s Levels of Moral Development .................... 26
Table 7. Taxonomy of Ethical Ideologies*.......................................................... 30
Table 8. PAVE Strategy*...................................................................................... 31
Table 9. Four components/questions for making a moral judgment*................. 32
Table 10. Aspects of Moral Judgment*............................................................... 34
Table 11. Means and Standard Deviations of DIT P-score and DIT2 N2-scores 58
Table 12. Mapping of SRM-SF items to sociomoral norms .......................... 70
Table 13. Chronological summary of moral reasoning measures ..................... 81
Table 14. 24 moral values in the TMV*............................................................... 85
Table 15. Invited schools and their responses .................................................. 91
Table 16. Scoring matrix of two-tier items....................................................... 98
Table 17. Original factor pattern........................................................................ 110
Table 18. Factor loadings with items removed .................................................. 112
Table 19. Goodness-of-fit indicators of estimation methods ............................ 115
Table 20. Scoring matrix of two-tier items with collapsed categories .......... 123
Table 21. Uncentralised item thresholds........................................................... 125
Table 22. Fit statistics for MRQ with collapsed categories .............................. 128
Table 23. DIF by level, gender, school and education stream ......................... 131
Table 24. Goodness-of-fit indicators of CFA with collapsed categories........140
Table 25. Test of measurement invariance by school level..........................144
Table 26. Test of measurement invariance by gender..................................144
Acknowledgements

I would like to thank the following people for their support in completing this work: (i) Dr. Elaine Chapman, my supervisor, for always encouraging and guiding me patiently; (ii) Professor David Andrich, my co-supervisor, for his valuable inputs; (iii) the Singapore Examinations and Assessment Board, for sponsoring this study; (iv) the Singapore Ministry of Education, for allowing me to collect data from schools; (v) the school leaders and teachers who supported and helped me in collecting data; (vi) my mother and my wife, for their unwavering support and help that sustained me through the challenges of doing an EdD and working full-time; (vii) my late father and my maternal grandmother, who encouraged me to pursue a higher degree before their passing; and (viii) my daughter, who inspired me to press on in challenging times.
Chapter 1. The Construct of Moral Reasoning

This chapter introduces the construct of moral reasoning, to provide a conceptual framework for the discussion of how moral reasoning is currently assessed in Chapter 2. Issues surrounding the definition of moral reasoning are first considered, followed by a summary of existing theoretical models of moral reasoning development. Research on how gender and culture may impact moral reasoning is then reviewed briefly, given that these attributes have been posited as potentially important moderating variables in the assessment of moral reasoning. The chapter concludes with a consideration of how moral reasoning has now been incorporated within the Singapore secondary curriculum, to provide a context for the research work presented within this thesis.

1.1. Issues in Defining Moral Reasoning

Paxton and Greene (2010) defined moral reasoning as a “conscious mental activity through which one evaluates a moral judgment for its (in)consistency with other moral commitments, where these commitments are to one or more moral principles and (in some cases) particular moral judgments” (p.516). Research interest in the development of moral reasoning or understanding within formal education was fuelled considerably by the appearance of Jean Piaget’s framework on the development of moral understanding, and then by the appearance of Lawrence Kohlberg’s stage-based model of moral reasoning (Palmer, 2011). Kohlberg’s model extended the notion of moral reasoning to ages well beyond those considered by Piaget, and is thus the model adopted most frequently in considering moral development at the secondary school level.
Since the appearance of these seminal theoretical frameworks, the education research literature has converged broadly on the notion that moral reasoning is the cognitive process that takes place before moral decisions or judgements are made. For example, Haidt (2001) stated that moral reasoning is the “conscious mental activity that consists of transforming given information about people in order to reach a moral judgment” (p.818).

As with other social constructs, however, the definition of moral reasoning remains highly contested. Saunders (2015) posed that many variations and interpretations of moral reasoning have arisen in large part due to a lack of clarity on the discourse of moral reasoning, and also, to the varied stipulations that have appeared to explain the notion of moral reasoning. The literature review undertaken in preparation for this research supported Saunders’ position. The literature on moral reasoning is characterised by a wide variety of terms and depictions, some connoting only slight variations in meanings, others highly disparate. Examples of terms that have been used at times interchangeably within the literature in relation to moral reasoning include moral orientation, moral intuition, moral intensity, ethical reasoning, ethical orientation, moral dimension, moral distinction, and moral judgment. In the next section, some key distinctions between the terms that have appeared in the research literature to date are drawn, to bring into sharper focus the construct of moral reasoning as the focus of the research work reported in this thesis.
1.1.1. Moral reasoning vs. ethics

Within the literature, there is considerable overlap in the concepts of moral reasoning and ethics. Indeed, research on ethics is often depicted as the study of cognitions that take place before moral decisions are made. Cheung (1999) opined that ethical reasoning facilitates ethical judgment and behaviour, and Roth (2005) posited that ethics may also be referred to as the study of value judgments. These are similar to O-Leary’s (2007) view that ethics is the study of “what we ought to do” (p.80) and that ethics address issues related to the guiding principles for humans to discriminate right from wrong. More recently, Ip (2013) stated that ethics is about “what should I do or how should I act” and that ethics involves understanding what is right or wrong, and what is good or bad and then “doing the right and good things” (p.106).

These descriptions underscore the significant overlap that exists between the constructs of ethics and moral reasoning. Indeed, Paxton and Greene’s (2010) definition of moral reasoning, presented above, bears a direct correspondence with some scholars’ depictions of how ethical reasoning results in the making of ethical judgments. For example, Vayrynen (2006) defined an ethical theory as one that “gives adequate moral guidance if it makes reliable strategies for acting well – for doing the right thing for the right reasons in particular situations – available to practical thinking” (p.293).
In an effort to differentiate ethics from morals, Keniston (1965) suggested that ethics are innate, while morals are learned codes of conduct. He used the term *ethics* to refer to an “individual’s thought-out, reflective and generalized sense of good and evil, the desirable and the undesirable, as integrated into his sense of himself and his view of the world” (p.628), while the term *morals* was used to describe context-specific rules. More recently, Baer (2005) also differentiated ethics from morality, when he described ethics as the study of moral choices and morality as the “system of rules and values governing those choices” (p.479).

Other scholars have suggested that moral reasoning and moral judgment can be encapsulated by the normative ethics categories suggested by Hurthhouse (2013) in the *Stanford Encyclopaedia of Philosophy*. These categories are: (i) *consequentialism* or *utilitarianism*, which emphasises the consequences of actions; (ii) *deontology*, which emphasises duties or rules; and (iii) *virtue ethics*, which suggest that an action of helping is considered positive regardless of the consequences. In this view, different forms of moral reasoning can be depicted using different categories of ethical reasoning. In a systematic attempt to map moral reasoning to ethical reasoning, Cheung (1999) used Kohlberg’s Theory of Moral Development and posited that each ethical reasoning paradigm is a level in the hierarchical Kohlbergian model. This mapping is presented in Table 1.
Table 1. Mapping of Ethical Reasoning to Kohlberg’s Stages*

<table>
<thead>
<tr>
<th>Ethical Reasoning Paradigm</th>
<th>Description</th>
<th>Kohlbergian Theory of Moral Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>Egoistic</td>
<td>Favours acquisition of interests for oneself</td>
<td>Pre-conventional level</td>
</tr>
<tr>
<td>Conventional</td>
<td>Refers to conformity to law, norms, and professional ethical codes</td>
<td>Pre-conventional level</td>
</tr>
<tr>
<td>Teleological</td>
<td>Emphasises maximization of benefits for society or the greatest number of people</td>
<td>Post-conventional level (Stage 5)</td>
</tr>
<tr>
<td>Deontological</td>
<td>Focuses on the intention for realizing important goals, ideals, and universally desirable values, including loyalty</td>
<td>Post-conventional level (Stage 6)</td>
</tr>
</tbody>
</table>

* Adapted from “Ethical judgment and ethical reasoning on business issues: A cross-lag model for university students In Hong Kong” by C.K. Cheung, 1999, College Student Journal, 33(4A).

1.1.2. Moral reasoning vs. moral intensity and moral orientation

In addition to the contested distinctions within the literature between the constructs of ethical and moral reasoning, there is potential for confusion between the notion of moral reasoning and other related constructs that have since been introduced. For example, Jones (1991) defined the concept of moral intensity as “the extent of issue-related moral imperative in a situation” (p. 372), and identified six attributes that would determine the level of moral imperative within any given situation: (i) the magnitude of consequences of a moral judgment; (ii) social consensus; (iii) the probability of the moral action producing benefit or harm; (iv) the temporal immediacy of the consequences; (v) the feeling of immediacy to those involved in the moral situation; and (vi) the concentration of effect. Based on the theoretical notions of moral reasoning discussed
previously, it is clear that some of the factors that would moderate levels of moral intensity, based on Jones’s model, are also key differentiating elements of many stage-based moral reasoning models (e.g., assessing the likelihood that the action will produce benefit or harm). Owing to commonalities of this kind, the distinction between these terms remains somewhat unclear.

Others have also attempted to differentiate between the terms *moral reasoning* and *moral orientation*. For example, in 2012, Levitt and Aligo introduced the notion of moral orientation as the perspective from which one approaches ethical decision making, and identified two main bases for such decisions: a *justice*-based orientation, which focuses primarily on factors such as equity and fairness; and a *care*-based orientation, which focuses more on meeting the needs of others. They asserted that their notion of moral orientation was grounded in Kohlberg’s, and later, Gilligan’s, works on moral development (discussed in more depth in later sections). Again, given this common theoretical grounding, there is a clear overlap in the notions of moral orientation and moral reasoning, although the notion of moral orientation is arguably more restricted.

Adding further to this debate, Dakin (2014) suggested moral orientation as a different way to look at moral reasoning, and argued that moral orientation could be seen as an individual's preferred moral reasoning style. While she did not provide distinctive definitions of moral reasoning and orientation, Dakin provided an example to underscore the difference between the two constructs. She suggested that someone at a particular moral reasoning stage based on Kohlberg’s model could still have a *preferred* orientation that is inconsistent with
this stage. Whilst Dakin’s arguments do suggest one basis on which the two terms can be differentiated, the integration offered is not definitive. In particular, these arguments leave many questions unresolved about the likely interplay between the two processes in determining an individual’s final moral or ethical judgements in any given situation.

In defence of retaining the construct of moral intensity, Lin, Tsou, Cho, Hsieh, Wu and Lin (2012) argued, in their study of medical students’ moral orientations, that moral orientation may be a more useful construct than moral reasoning in exploring decision-making in medicine. They stated that the moral decision-making processes of clinicians will be moderated by three main factors: (i) the doctor's level of clinical and ethical knowledge; (ii) the doctor’s moral reasoning skills; and (iii) the doctor’s attitudes. Importantly, however, moral orientation was not mentioned within any of these components. Furthermore, the arguments put by Lin et al. do not consider how the notion of moral intensity relates to the four primary ethical principles that medical practitioners are obliged legally to operate within: beneficence (i.e., acting such that the patient will always benefit from that action); non-maleficence (i.e., not harming the patient in whatever actions taken); respect for autonomy (i.e., holding the patient’s decision in the highest regard and avoiding deception in dealing with patients); and justice (i.e., acting in a manner such that all patients are treated equally and fairly regardless of relationships) (Gillion, 1994). However, each of these principles can be mapped to stages within Kohlberg’s Theory of Moral Development. Given this, it is unclear at this point how the notion of moral orientation is a more useful
construct than that of moral reasoning in considering decision-making processes in medical practitioners.

1.1.3. Definition adopted within the present study

The definitions and discussions presented above indicate that, while there are debates on whether ethics are morals, whether ethical reasoning is moral reasoning, or whether moral orientation is moral reasoning, there has been some convergence in the literature on the notion that moral (or ethical) reasoning is a cognitive process that conciliates moral (or ethical) disequilibrium in an individual, and eventually prompts the moral or ethical decision made. Also, as noted, while some scholars have attempted to differentiate ethics from morals, these terms have also come to be used interchangeably in the literature by various others. Based on a thorough review of the literature within the area, it was decided that, in the context of this thesis, the term moral reasoning would be used simply to refer to the thought processes undertaken by an individual in making a judgment that requires consideration of moral factors.

1.2. Theoretical Frameworks on Moral Reasoning

Many scholars have proposed that the development of moral reasoning proceeds through a predictable sequence of stages, and that individual moral responses can thus be mapped to a given developmental stage. This section presents some of the most prominent theoretical frameworks on the development of moral reasoning that have appeared to date within the education research literature.
1.2.1. Piaget’s theoretical framework on moral understanding

Jean Piaget is generally acknowledged to have pioneered the construction of formal theoretical models on the development of moral reasoning in children. His most important work in the area first appeared in the book *The Moral Judgment of the Child*, published in 1932. Snarey and Samuelson (2008) stated that, in Piaget’s view, every child is a ‘moral philosopher’, and will always be constructing actively his or her notions of what is right and wrong whilst interacting with the world. This work stemmed from Piaget’s original research with young children, in which he concluded, based on his observations, that “young children tend to conceptualize morality in terms of obedience to adults, whereas older children tend to conceptualize it in terms of cooperation with peers” (Krebs & Denton, 2005, p.629).

Patanella (2011) stated that, according to Piaget, children go through three broad phases in developing their moral understandings. In Phase 1 (zero to four years old), the child is unconcerned about morality, as he or she is still trying to master motor and social skills. In Phase 2 (seven to eleven years old), the child exhibits unconditional respect for rules and submission to authority (heteronomous morality). In Phase 3 (eleven years or older), the child recognises that rules are arbitrary and may be altered with group consensus. In the latter phase, the child will also come to recognize that the intentions, rather than merely the consequences, of an action should be considered in judging the morality of that action. In establishing these three phases, Piaget theorised that this difference reflected the fact that older children had greater cognitive ability to appreciate various different perspectives, including reciprocal relationships with peers.
Phase 1, in which a child is unconcerned about morality, is not commonly mentioned and hence, Piaget’s theory is more often effectively a two- rather than three-phase theory. Piaget’s two-phase theory is detailed below in Table 2. As indicated, based on this model, children move from what is termed heteronomous to autonomous morality as they develop. It should be noted that Piaget did not originally use the term ‘stages’ in describing his theoretical model. He advocated instead the term phases, recognising that under different circumstances, children can exhibit elements of both heteronomous and autonomous morality.

Table 2. Piaget’s Two-phase theory*

<table>
<thead>
<tr>
<th>Cognitive development of morality</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Heteronomous morality explanation</td>
<td>Autonomous morality explanation</td>
</tr>
<tr>
<td>Morality is based on unilateral respect for authorities and the rules they prescribe. Fairness is understood as obedience to authorities and conformity to these rules. Consequences are understood as concrete objective damage and is more relevant than intention. Expiatory punishment is the favoured way to make things right.</td>
<td>Morality is based on respect, reciprocity and equality among peers. Fairness is a mutually agreed cooperation and reciprocal exchange. Intention and consequence can be kept in mind concurrently. Punishment by reciprocity is favoured.</td>
</tr>
</tbody>
</table>

* Adapted from Snarey and Samuelson (2008, p.55)

Whilst Piaget’s work is recognised to have pioneered formal theories on the development of moral reasoning, various criticisms on Piaget’s theory have appeared. According to Lourenco and Machado (1996), ten criticisms of Piaget’s cognitive theory approach to moral reasoning have been particularly prominent:

(i) the underestimation of the competence of children implicit within the model;
(ii) the failure of empirical studies to establish age norms commensurate with the model;

(iii) a proposed negative characterisation of development within the model;

(iv) a proposed 'extreme competence' assumption underlying the approach;

(v) the model’s neglect of social factors that may influence moral understandings within children;

(vi) the lack of corroboration of the model’s predictions from empirical studies;

(vii) the limited explanation of the model proffered by Piaget;

(viii) the proposed paradoxical nature of Piaget’s approach, given that the model relies upon assessing thinking through language;

(ix) the omission of post-adolescence development within the model; and

(x) the notion that Piaget’s approach’s appeals to inappropriate models on logic.

In his own empirical work, Piaget relied heavily on qualitative research methods (e.g., clinical interviews and observations) to establish his theory of moral development. Hence, it is not surprising that some of the strongest critiques of his model have commented upon a lack of extensive empirical evidence for the model. Since Piaget’s research method relied on interviews and observations, social factors (e.g., social-economic status, and culture) were also neglected. Patanella (2011) stated that these criticisms have led to questions regarding the universality of the ages (0-4; 7-11; > 11) that Piaget proposed in connection with his theory.

Empirical studies have produced partial support for Piaget’s theory at best. For example, Durkin (1959a), in his study with three groups of a total of 101 boys and
girls (mean age 7.8, 10.9, and 13.9), found some support for Piaget’s notion that justice concepts are related to age, but no support for Piaget’s more specific view that children accepted reciprocity as a concept of justice as they aged. In a second study by the same researcher (Durkin, 1959b) with 119 boys and girls, no results were produced to support Piaget’s theory that older children tended more towards autonomous morality than do younger children. Durkin found that the oldest of the children in his research, as well as the younger ones, tended to seek justice in an authority person (heteronomous morality). This is inconsistent with Piaget’s theory that reciprocity as a concept of justice increases with age.

In a later study with 78 boys and 87 girls ranging in age from 5 to 11 years, Bandura and McDonald (1963) found that while subjective moral orientations generally increased with age, Piaget's theory of demarcated sequential stages of moral development could not be substantiated. In this latter study, children at all age levels exhibited discriminative repertories of moral judgments concurrently. Similar results were obtained in a replication of this study by Cowan, Langer, Heavenrich and Nathanson (1969). Cowan et al. concluded that their replication, as well as providing support for Bandura and McDonald’s earlier findings, also provided evidence that children’s moral responses can be modified in either developmental direction by adults’ influences, a factor that is not accounted for within Piaget’s developmental model.

In defence of Piaget’s original model of moral understanding, it should be noted here that Piaget explicitly preferred the term ‘phases’ to the term ‘stages’ in describing his categories of moral understanding. He acknowledged that
children could sometimes exhibit both types of moral understanding (heteronomous and autonomous) under different circumstances. In light of this, some degree of caution is warranted in the interpretation of findings that purport to contradict or refute Piaget’s original theory. Irrespective of the various conceptual, ideological, and other limitations that have now been highlighted with regard to Piaget’s early work in moral reasoning, Piaget’s model is recognised to have laid the groundwork for subsequent theorists (such as Lawrence Kohlberg) to develop more elaborate models, which have been found to align more consistently with empirical developmental data.

1.2.2. Kohlberg’s Stage Theory of Moral Reasoning

Lawrence Kohlberg’s (1984) work on moral reasoning followed directly from Piaget’s earlier model, providing a more specific and elaborate framework on how moral reasoning develops across the lifespan. Kohlberg defined stages of moral reasoning by the ways in which individuals reasoned around and justified their behaviour (Palmer, 2011). Kohlberg first developed his Theory of Moral Development by observing how children responded to moral dilemmas. He presented the dilemmas as short stories, and asked probing questions to elicit how the children felt about the actions of the characters in the dilemmas.

Based on his observations, Kohlberg concluded that there were three broad developmental levels of moral reasoning, which could each be further subdivided into two distinct stages. Each of the two stages within the three broad levels of moral reasoning consists of one stage that is similar in orientation to, but less complex, than the other, and one stage that is more advanced from a
sociomoral point of view. The key elements of Kohlberg’s theoretical model of moral development are summarised in Table 3.

In constructing his theory, Kohlberg asserted that moral reasoning arises from assuming a deontic stand or prescriptive judgment of obligations and corresponding rights, and that moral judgment is generally anchored on one of four moral orientations: (i) the impartial following of rules and normative roles; (ii) utilitarianism, in which the welfare of each person in a moral situation is maximised; (iii) the idealistic view of seeking harmony or integrity of one or a social group; or (iv) the maintenance of fairness, balance of perspectives, equity, and social contract (Colby, Kohlberg, Gibbs & Lieberman, 1983).

Kohlberg’s proposed stages are hierarchical, such that each may be considered a more sophisticated response to moral dilemmas than the previous (Kohlberg, 1984). According to Krebs and Denton (2005), Kohlberg asserted that “people interpret all of the moral information they process in terms of the structures of moral reasoning that define their current stage of moral development” (p.630). Essentially, Kohlberg’s theory implies that as individuals mature socio-cognitively, they will develop a more complex and inclusive view of justice, and hence, of moral reasoning and judgment. As a result, their moral reasoning and judgment will progress in an upward direction without the skipping or mixing of stages, though the application of reasoning in adjacent stages is conceivable. As such, Kohlberg’s theory stipulates that stages are reached in an invariant sequence, and that an individual cannot regress to an earlier stage.
Table 3. Summary of Kohlberg’s Six Stages of Moral Development*

<table>
<thead>
<tr>
<th>Level</th>
<th>Stage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1 (Pre-Conventional): Most common in children (up to age nine), although adolescents or adults may also reason in this way. At this level, the morality of an action is judged by its direct consequences.</td>
<td>1. Obedience and punishment orientation</td>
<td>Individuals focus on the direct consequences of their actions on themselves.</td>
</tr>
<tr>
<td></td>
<td>2. Self-interest orientation</td>
<td>Behaviour is defined by whatever the individual believes to be in their best interest; there is an interest in the needs of others, but only insofar as serving these might further the individual's own interests.</td>
</tr>
<tr>
<td>Level 2 (Conventional): Typical of adolescents (up to age 20) and some adults. At this level, the morality of actions is judged by comparing them to societal views and expectations.</td>
<td>3. Interpersonal accord and conformity</td>
<td>Focus on approval or disapproval from others, based on the notion that being regarded as good benefits the self - morality of an action is often judged by evaluating its consequences in terms of a person's relationships.</td>
</tr>
<tr>
<td></td>
<td>4. Authority and social-order maintaining orientation</td>
<td>Laws, dictums and social conventions are valued because of their importance in maintaining an effective society.</td>
</tr>
<tr>
<td>Level 3 (Post-Conventional): Most common in adulthood, though many adults do not reach this stage. At this level, individuals live by their own ethical principles — principles that typically include such basic human rights as life, liberty, and justice.</td>
<td>5. Social contract orientation</td>
<td>Laws are regarded as social contracts, which reflect different opinions, rights and values, which should be mutually respected as unique to each person or community.</td>
</tr>
<tr>
<td></td>
<td>6. Universal ethical principles</td>
<td>Moral reasoning is based on application of universal ethical principles. Laws are valid only insofar as they are grounded in justice, and a commitment to justice carries with it an obligation to disobey unjust laws.</td>
</tr>
</tbody>
</table>

As indicated in Table 3, Kohlberg classified his six stages of moral development into three main development levels of morality: *pre-conventional, conventional* and *post-conventional*. Kohlberg also asserted that the stages in his theory are behaviourally non-prescriptive, in that each stage is defined by the *form* of reasoning adopted by an individual, and not by the *content*, or by any specific decisions that the individual reaches. In an elaboration of this viewpoint, Bandura (1991) explained that Kohlberg’s stages were developmental, and determine *reasons* for potential actions, but not *what* actions should be taken.

Whilst Kohlberg’s theory has had a profound effect on the way that moral reasoning has been conceptualised within the education research literature, the theory has not been without its critics. One criticism has arisen from the fact that Kohlberg himself, in his own longitudinal research work, found that it was not tenable to retain the highest level of the model (Stage 6), because so few individuals ever actually attained this level. This action has caused some to doubt the universality of Kohlberg’s theory (e.g., Green & Snarey, 2011). In defence of Kohlberg’s approach, some researchers, such as Gibbs, Widaman and Colby (1982) argued that the empirical rarity of individuals exhibiting Stage 6 reasoning within one particular study does not provide warrant for questioning the utility of that stage, or of the model as a whole, across all samples. Based on arguments of this kind, Stage 6 has been retained in the theoretical model, although it is widely recognised that many will never attain this level of moral reasoning.

Other criticisms have arisen with respect to the prescriptive nature of the stages within the model, and the assumption that development must proceed...
invariantly according to those stages (e.g., Donenberg & Hoffman, 1988). A small number of studies have produced results that challenge Kohlberg’s notion of invariant stages in the development of moral reasoning. For example, Holstein (1976), in his study with 53 13 year old subjects, found that regression in moral reasoning could occur at the higher stages within the model, thus challenging the irreversibility of Kohlberg’s stage sequence. In contrast, many other studies have produced evidence in support of Kohlberg’s invariant stage theory. These studies, in general, have reported progression of moral reasoning in a sequential nature that is aligned with Kohlberg’s model across different cultures and contexts (Page, 1981; Walker, de Vries, & Bichard, 1984). In one particularly influential paper on this topic, Snarey (1985) reported on a meta-analysis of 45 studies across 27 countries, which suggested that Kohlberg’s invariant stage sequence was well-supported. Based on this same analysis, Snarey noted that regressions were rare, and could be attributed to measurement error in those cases in which it was observed.

Other scholars have challenged the typical age ranges suggested by Kohlberg for development through each of the model’s stages, citing empirical evidence that contradicts these assumptions. For example, in a study by Smetana and Braeges (1990), 108 toddlers (equal number of 26-, 34- and 42-month boys and girls) were interviewed on how they would respond to moral situations (e.g., a child is hitting another child. Is it bad for this child to hit another child? Is it still wrong for this child to hit another child if his/her teacher did not tell him/her it is wrong to do so?). Smetana and Braeges found that while the 26-month-old toddlers were not sensitised to the difference between moral or social convention, the 34-
month-old toddlers were already able to judge that moral transgressions were more wrong than social conventional transgressions.

In another study, Melanie et al. (2001) interviewed 72 children (35 4.5 year olds and 37 5.5 year olds, nearly equally divided by gender) from European American (71%), Asian American (17%), and African American (12%) middle-class backgrounds individually. They presented the children with two gender-biased activities (doll playing and truck playing) and the children had to select either a girl or boy to join a group of children of the same gender for the play activity, on the pretext that there was only enough space for one more child to join. Contrary to the conventional thinking that the children would select automatically based on gender, the researchers found that fairness had a greater influence on the children's decisions (i.e., the children provided a variety of moral reasons related to prosocial behaviour, fairness and equality). This study suggests that even children at four years of age (typically, the pre-conventional level) can appreciate the concept of fairness. Hence, in this regard, it appears that there are refinement possibilities to Kohlberg’s or even Piaget’s stipulated age ranges for each level of moral reasoning. More specifically, individuals expected to have a pre-conventional level of moral reasoning may already be beyond that level.

Various scholars have also pointed to methodological issues in Kohlberg’s original research, which bring into question some aspects of the model that was then developed. For example, Carol Gilligan, Kohlberg’s research assistant, took issue with how Kohlberg drew conclusions based on his interviews with an all-male sample, citing various reasons why males and females might progress
through different moral reasoning stages. Gilligan’s work, in which she proposed alternative levels of moral development that better accommodate how females might respond to Kohlberg’s moral dilemmas, are considered in a later section.

1.2.3. Frameworks by Rest and colleagues

As noted above, one concern that has been raised with the Kohlberg model relates to the empirical rarity of individuals attaining the highest stage within the model. In response to such criticisms, Rest and colleagues (Rest, 1979; Rest, Naravaez, Bebeau & Thoma, 1999a) reconceptualised Stage 6 of Kohlberg’s model, and proposed a more broadly defined post-conventional stage equivalent for this stage. In their alternative to the Kohlberg model (the Schema Theory), Rest et al. attempted to encompass an individual’s shift from conventional to post-conventional thought based on socially constructed norms. Unlike Kohlberg, Rest and colleagues did not attempt to postulate that the equivalent of the post-conventional moral reasoning level in the Schema Theory was universally applicable. On the contrary, they stated that “moral obligations are to be based on shared ideals, which are reciprocal and are open to debate and tests of logical consistency, and on the experience of the community” (Rest et al., 1999a, p. 307).

In reformulating his 1969 dissertation, Rest (1979) proposed moral reasoning as a function of two factors: (i) concepts of how expectations of others and self are coordinated or shared, and (ii) concepts of how the equilibrium of interests between cooperating individuals in a situation can be achieved. He mapped these two factors to Kohlberg’s progressive stages and posited that individuals ascribe to moral reasons and hence judgments based on these two factors (see Figure 1).
Essentially, Rest equates moral reasoning with the way in which social justice is achieved through the equilibration of interests and how each individual ascribes to a corresponding level of moral development and hence provides cooperation to resolve a moral dilemma. In an attempt to detail Kohlberg’s stages, Rest (1979) suggested action statements characteristic of each stage. Table 4 presents these characteristic statements.

Rest et al. (1999a) later proposed three developmental schemas as an alternative to Kohlberg’s six stages, based on a factor analysis of the responses for the Defining Issues Test (DIT), an instrument that seeks to obtain moral judgment data. The Schema Theory includes: (i) the personal interest schema, derived from Kohlberg’s stages two and three, in which the individual considers only what he or she will gain or lose in a moral dilemma and how others will like him based on his judgment; (ii) the maintaining norms schema, derived from Kohlberg’s stage four, in which the individual seeks to maintain law and the established social order; and (iii) the post-conventional schema, derived from Kohlberg’s stages five and six, in which the individual considers moral ideals and obligations (i.e., human rights) that are fully reciprocal and debatable in deciding upon an appropriate action to take when confronted with a moral dilemma.
<table>
<thead>
<tr>
<th>Stage</th>
<th>Coordination of expectations about actions (how rules are known and shared)</th>
<th>Schemes of balancing interests (how equilibrium is achieved)</th>
<th>Central concept for determining moral rights and responsibilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1</td>
<td>The caretaker makes known certain demands on the child's behavior.</td>
<td>The child does not share in making rules, but understands that obedience will bring freedom from punishment.</td>
<td>The morality of obedience: “Do what you’re told.”</td>
</tr>
<tr>
<td>Stage 2</td>
<td>Although each person is understood to have his own interests, an exchange of favors might be mutually decided.</td>
<td>If each party sees something to gain in an exchange, then both want to reciprocate.</td>
<td>The morality of instrumental egoism and simple exchange: “Let’s make a deal.”</td>
</tr>
<tr>
<td>Stage 3</td>
<td>Through reciprocal role taking, individuals attain a mutual understanding about each other and the ongoing pattern of their interactions.</td>
<td>Friendship relationships establish a stabilized and enduring scheme of cooperation. Each party anticipates the feelings, needs, and wants of the other and acts in the other’s welfare.</td>
<td>The morality of interpersonal concordance: “Be considerate, nice, and kind, and you’ll get along with people.”</td>
</tr>
<tr>
<td>Stage 4</td>
<td>All members of society know what is expected of them through public institutionalized law.</td>
<td>Unless a society-wide system of cooperation is established and stabilized, no individual can really make plans. Each person should follow the law and do his particular job, anticipating that other people will also fulfill their responsibilities.</td>
<td>The morality of law and duty to the social order: “Everyone in society is obligated and protected by the law.”</td>
</tr>
<tr>
<td>Stage 5</td>
<td>Formal procedures are institutionalized for making laws, which one anticipates rational people would accept.</td>
<td>Law-making procedures are devised so that they reflect the general will of the people, at the same time insuring certain basic rights to all. With each person having a say in the decision process, each will see that his interests are maximized while at the same time having a basis for making claims on other people.</td>
<td>The morality of societal consensus: “You are obligated by whatever arrangements are agreed to by due process procedures.”</td>
</tr>
<tr>
<td>Stage 6</td>
<td>The logical requirements of non-arbitrary cooperation among rational, equal, and impartial people are taken as ideal criteria for social organization which one anticipates rational people would accept.</td>
<td>A scheme of cooperation that negates or neutralizes all arbitrary distribution of rights and responsibilities is the most equilibrated, for such system is maximizing the simultaneous benefit to each member so that any deviation from these rules would advantage some members at the expense of others.</td>
<td>The morality of non-arbitrary social cooperation: “How rational and impartial people would organize cooperation is moral.”</td>
</tr>
</tbody>
</table>

Figure 1. Mapping of moral reasoning to its two functions*  

<table>
<thead>
<tr>
<th>Kohlberg’s stage of moral development</th>
<th>Characteristic statements by Rest (1979)</th>
</tr>
</thead>
</table>
| **1. Obedience and punishment orientation** | a) Right and wrong are defined in terms of obedience to fixed rules.  
b) Punishment follows disobedience and one who is punished is bad. |
| **2. Self-interest orientation** | a) An act is right if it serves one’s interest and desire.  
b) One should obey the law if it is practical to do so.  
c) Cooperative interaction is based on simple exchange. |
| **3. Interpersonal accord and conformity** | a) An act is considered good if the motivation is pro-social.  
b) An act of being moral hinges on whether it is approved by others. |
| **4. Authority and social-order maintaining orientation** | a) An act is morally right if it maintains social order and is expected by society.  
b) Values are derived from and subordinated to the social order and maintenance of law i.e., an act can be guided by values; values are a subset and are consistent with established laws that maintain social order. Personal values cannot supersede the law.  
c) Respect for delegated authority is part of one’s obligations to society. |
| **5. Social contract orientation** | a) Moral obligation derives from voluntary commitments of society’s members to cooperate i.e., an act is moral if societal law agrees and one is part of this society voluntarily.  
b) Procedures exist for selecting laws that maximise welfare as discerned in the majority will.  
c) Basic rights are preconditions to social obligations. |
| **6. Universal ethical principles** | a) Moral judgments are ultimately justified by principles of ideal cooperation.  
b) Individuals each have an equal claim to benefit from the governing principles of cooperation. |
Rest et al. (1999a) stated that the Schema Theory differs from Kohlberg’s in that all three moral schemas are more concrete and specific than Kohlberg’s stages. The schemas are also interconnected and developmental, and hence allow for more variability in moral development than Kohlberg’s invariant sequential stages. Given this ‘variable’ nature of the Schema Theory, which permits the non-linear movement of moral development, this theory has attracted fewer critics than did Kohlberg’s theory.

Despite the arguments posed by Rest et al. in support of the Schema Theory, this model is highly similar to Kohlberg’s approach to conceptualising moral development. Table 5 presents a mapping of Kohlberg’s Theory and the Schema Theory. It should be noted here that the Schema Theory postulated by Rest et al. (1999a) did not include stage one of Kohlberg’s Theory, because Rest and colleagues argued that this theory was designed to capture moral reasoning development in people with at least a 12-year-old reading level. Nonetheless, Walker (2002) commented that the omission of stage one in the Schema Theory is a major limitation as “considerable development and moralisation take place” during childhood (p.362).
Table 5. Mapping of Kohlberg’s Theory and the Schema Theory

<table>
<thead>
<tr>
<th>Kohlberg’s Theory of Moral Development (stages and levels)</th>
<th>Schema Theory by Rest et al. (1999a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1 - Obedience and punishment orientation</td>
<td>Level 1 – Pre-conventional</td>
</tr>
<tr>
<td>Stage 2 - Self-interest orientation</td>
<td>Level 1 – Pre-conventional</td>
</tr>
<tr>
<td>Stage 3 - Interpersonal accord and conformity</td>
<td>Level 2 – Conventional</td>
</tr>
<tr>
<td>Stage 4 - Authority and social-order maintaining orientation</td>
<td>Level 2 – Conventional</td>
</tr>
<tr>
<td>Stage 5 - Social contract orientation</td>
<td>Level 3 – Post-conventional</td>
</tr>
<tr>
<td>Stage 6 - Universal ethical principles</td>
<td>Level 3 – Post-conventional</td>
</tr>
</tbody>
</table>

*Numbers represent corresponding stages from Kohlberg’s theory.

1.2.4. Theoretical framework by Gilligan

As noted previously, Carol Gilligan was one of the earliest critics of Kohlberg’s stage theory of moral development. Gilligan based much of her criticism on limitations in the empirical methods that Kohlberg used in deriving his stages, and in particular, on the fact that Kohlberg’s sample was exclusively male. Following her critique of Kohlberg’s model as male-biased, Gilligan (1982) proposed an alternative model that she posed would be more suitable for determining the moral reasoning stages of female participants. She asserted that males follow Kohlberg’s more *justice*-oriented model, while females follow one that is more *care*-oriented. She also opined, however, that both models should be seen as developmental, not in competition, and not mutually exclusive. Gilligan was of the opinion that females might be disadvantaged based on Kohlberg’s model. For example, she wrote:
Prominent among those who thus appear to be deficient in moral development when measured by Kohlberg’s scale are women, whose judgments seem to exemplify the third stage of his six-stage sequence. At this stage morality is conceived in interpersonal terms and goodness is equated with helping and pleasing others. This conception of goodness is considered by Kohlberg and Kramer (1969) to be functional in the lives of mature women insofar as their lives take place in the home. Kohlberg and Kramer imply that only if women enter the traditional area of male activities will they recognize the inadequacy of this moral perspective and progress like men towards higher stages where relationships are subordinated to rules (stage four) and rules to universal principles of justice (stages five and six) (Gilligan, 1982, p. 18).

Gilligan (1982) argued that females who possess the very traits considered by society to be ‘good’ in women (e.g., their care for others) will penalised by Kohlberg’s model and deemed as deficient in moral reasoning and development. Hence, she proposed a more encompassing alternative. In her three-level framework, females progress through levels via an increased sense of self, rather than strictly through advancing cognitive ability. Level three of both Kohlberg’s and Gilligan’s models are the most distinct, in that Kohlberg’s level focusses on justice, while Gilligan’s focus is on balancing between care for self and others without any explicit consideration of justice. Nonetheless, level one of both Kohlberg’s and Gilligan’s frameworks are comparable. A comparison of the frameworks by Kohlberg and Gilligan is presented in Table 6.
Table 6. Kohlberg’s and Gilligan’s Levels of Moral Development

<table>
<thead>
<tr>
<th>Level</th>
<th>Kohlberg’s Levels of Moral Development</th>
<th>Gilligan’s Levels of Moral Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Pre-conventional (deferring to authority and self-interest)</td>
<td>Care for self (individual survival)</td>
</tr>
<tr>
<td>Level 2</td>
<td>Conventional (being bound by duty and maintaining law or social norms)</td>
<td>Care for others (self-sacrifice)</td>
</tr>
<tr>
<td>Level 3</td>
<td>Post-conventional (universal principles of justice; moral ideals and obligations are fully reciprocal and debatable)</td>
<td>Balance between care for self and others</td>
</tr>
</tbody>
</table>

As with Kohlberg’s model, Gilligan’s model has been subjected to various criticisms within the literature. In particular, it appears that Gilligan’s assertion that males tended towards the justice orientation while females tended towards the care orientation was unsupported. For example, in an attempt to evaluate Gilligan’s assertion of gender bias in Kohlberg’s model, Walker (1984) conducted a meta-analysis of studies across varied age groups and concluded that sex differences in moral reasoning were non-significant. In this analysis, the few studies that indicated statistically significant differences between males and females were confounded by extraneous variables such as level of education and occupation.

More recent research has supported findings reported earlier by Walker (1984). For example, Pratt, Golding, Hunter and Sampson (1988) reported that, based on a study of 72 subjects across three age groups (i.e., 18 to 24; 30 to 45; 60 to 75), while women tended more than men towards Gilligan’s care orientation, gender differences were not as pervasive as Gilligan argued. On the contrary, the age and stage level of the participant and his/her education level had a much greater
influence on his or her moral reasoning level and style. The finding that gender differences were not as pervasive as Gilligan had argued has also been reported in two more recent studies. In their study, Crandall, Tsang, Goldman and Pennington (1999) used a real-life dilemma to ascertain whether gender differences in moral reasoning based on empirical findings were valid. Using a sample of 242 undergraduates (100 males, mean age = 18.0 years) from an introductory psychology class at the University of Florida, they found that while gender differences were apparent within moral dilemmas as hypothesised (i.e., males tended towards a justice orientation and females tended towards a care orientation), moral orientations were dilemma dependent, and both males and females were flexible in their use of justice and care orientations. This presents a challenge to Gilligan’s assertion that females would automatically be disadvantaged based on Kohlberg’s model.

The results of two major reviews of the field have generally supported the findings reported in the studies summarised above. In one early review of gender differences in moral development and acquisition by Woods (1996), no significant difference was observed across studies in the way males and females make moral decisions. Also, more recently, Jaffee and Hyde (2000) concluded from their meta-analysis of studies on gender differences in moral reasoning, that findings “do not offer strong support for the claim that the care orientation is used predominantly by women and that the justice orientation is used predominantly by men” (p.703). Thus, whilst Gilligan’s cautions with respect to Kohlberg’s empirical research appear to have been well-founded, subsequent
research in which males and females have been compared in terms of their moral reasoning approaches have not supported the notion that Kohlberg’s use of an exclusive male sample resulted in a fundamentally gender-biased model.

1.2.5. Alternative frameworks and strategies

Various other scholars have attempted to establish frameworks or models alternative to Kohlberg’s and its derivatives. For example, Forsyth (1980) suggested that while Kohlberg’s model and some of its derivatives are equally valid, they could all be encompassed well by her Taxonomy of Ethical Ideologies, constructed as part of developing her Ethics Position Questionnaire (EPQ). Forsyth’s (1980) EPQ comprises two subscales: (i) the idealism subscale, which assesses an individual’s degree of agreement with items such as “if an action could harm an innocent other, then it should not be done” and “it is never necessary to sacrifice the welfare of others”; and (ii) the relativism subscale, which assesses an individual’s degree of agreement with items like “there are no ethical principles that are so important that they should be part of any code of ethics” and “whether a lie is judged to be moral or immoral depends upon the circumstances surrounding the action”. Essentially, individuals who score high on the idealism subscale of the EPQ would generally subscribe to the notion that positive consequences can always be obtained. Individuals who score high on the realism subscale would be those who believe that absolute and universal moral rules are not useful when making moral judgments.
Based on these two subscales, Forsyth proposed four categories of individuals that span between totally abiding and totally rejecting universal moral rules in moral situations (see Table 7). The four categories are: (i) situationists, who seek the best outcomes of their decisions in place of just pure considerations of moral rules; (ii) absolutists, who are similar to situationists though they think that all actions should be consistent with absolute moral principles; (iii) subjectivists, who are similar to situationists but they think it is impossible to avoid negative behaviours and hence describe all moral actions as subjective rather than absolute; and (iv) exceptionists, who feel that there are general moral guidelines, though in certain instances, these moral guidelines may not be adhered to avoid negative consequences. These four proposed quadrants share significant commonalities with Kohlberg’s and Gilligan’s levels of moral development, in that situationists, subjectivists and exceptionists in essence reflect a balance between a justice orientation and care for self or others when making moral decisions. The absolutist perspective, which reflect an adherence to the view that moral actions must be followed regardless of the consequences, can also be seen as a radical adherence to a justice orientation.
Table 7. Taxonomy of Ethical Ideologies*

<table>
<thead>
<tr>
<th></th>
<th>Relativism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idealism</td>
<td>High</td>
</tr>
<tr>
<td>High</td>
<td>Situationists Rejects moral rules; advocates individualistic analysis of each act in each situation; relativistic</td>
</tr>
<tr>
<td>Low</td>
<td>Subjectivists Appraisals based on personal values and perspective rather than universal moral principles; relativistic</td>
</tr>
<tr>
<td></td>
<td>Exceptionists Moral absolutes guide judgments by pragmatically open to exceptions to these standards; utilitarian</td>
</tr>
</tbody>
</table>

* Adapted from Forsyth (1980, p.176)

Beyond theories, various strategies have also been posed in the literature on approaches that individuals can use to address moral dilemmas. Strategies are critical to the development of instruments measuring moral reasoning, in particular, for item development in that they provide information on how individuals might respond to moral dilemmas. Henderson (2005), for example, put forward the Principles, Agreements, Virtues and End consequences (PAVE) moral reasoning strategy as a means by which moral reasoning could be conceptualised. The PAVE strategy is a linear framework that leads an individual from a moral dilemma to a possible outcome. The PAVE strategy appears to be applicable across different ages as long as appropriate contexts are considered. For example, younger children need more information in dealing with a moral dilemma while older children should be encouraged to probe more instead of spoon-feeding them with information. Table 8 depicts this strategy.
Table 8. PAVE Strategy*

<table>
<thead>
<tr>
<th>Explanation</th>
<th>Principles</th>
<th>Agreements</th>
<th>Virtues</th>
<th>End consequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>What makes an action right is whether it upholds a certain principle. Principles are like duties or rules that apply to any set of circumstances. So when someone argues that it is never right to murder, they are thinking about principles. Principles are very useful for dealing with large groups of people. Many laws are based on principles. Some common principles are: Do no harm. Always tell the truth. Keep your promises. Be fair.</td>
<td>What makes an action right is whether it is consistent with what people involved expect to happen. People willingly enter into contracts with each other, either formally or informally, written or spoken, about the way they agree to treat each other.</td>
<td>What makes an action right is whether it is what a virtuous person would do. Virtues are character traits, like courage, compassion and integrity. A virtuous person will always do the right thing out of habit because it is in their character to do so.</td>
<td>What makes an action right is whether it has good consequences, that is, whether it increases the welfare of the people affected by it. By ‘good’ we might mean happiness, well-being, pleasure, interest or satisfaction. If large numbers of people are affected, you could consider the greatest good for the greatest number.</td>
<td></td>
</tr>
</tbody>
</table>

* Adapted from Henderson (2005, p.185)
Henderson (2005) posited that when faced with moral dilemmas, it is assumed that individuals generally want to do the right thing. He proposed four questions that an individual could ask in leading up to a possible outcome: (i) What could I do that is morally right?; (ii) What should I do in this situation?; (iii) What will I do?; and (iv) Why will I choose that course of action? These four questions share much in common with the more established four components proposed by Rest (1979) and Wilson (1967). Table 9 presents the four components, and suggests common strands amongst these models, to demonstrate that while there are three distinct four component models, similarities exists amongst the four.

Table 9. Four components/questions for making a moral judgment*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>What could I do that is morally right?</td>
<td>moral sensitivity</td>
<td>Sensitising and identifying oneself with others</td>
<td>Exploring possible actions by sensitising oneself with others</td>
</tr>
<tr>
<td>What should I do in this situation?</td>
<td>moral judgment</td>
<td>Insight into the feelings of others and self</td>
<td>Exploring how possible judgments would impact others’ emotions</td>
</tr>
<tr>
<td>What will I do?</td>
<td>moral motivation</td>
<td>Knowledge of other relevant facts</td>
<td>Prioritising what to do based on self-motivation, values and knowing all facts within reach</td>
</tr>
<tr>
<td>Why will I choose that course of action?</td>
<td>moral character</td>
<td>Ability to decide and act on it</td>
<td>Appreciating the consequences of the action and having the courage to acting on it</td>
</tr>
</tbody>
</table>

* As proposed by Henderson (2005), Rest (1979) and Wilson (1967)

1.3. Culture and Moral Reasoning

One factor that is frequently cited as a potential moderating variable in determining individuals’ moral reasoning development is culture. Culture is
generally seen as multi-faceted, and no single definition has been accepted universally which encompasses all of these facets. Williams (1983), an influential figure in the field of cultural studies, stated that culture “is one of the two or three most complicated words in the English language” (p.87). Nonetheless, culture can be contextualised and hence defined based on the desired level of granularity. At a macro level, culture can be explained by a belief, thinking or behaviour widely accepted or even encouraged by a society.

Based on longitudinal research studies he was involved, Kohlberg claimed that moral reasoning is culturally independent and that there are universally moral concepts and values or principles. He wrote:

almost all individuals in all cultures use the same thirty basic moral categories, concepts, or principles; and (b) all individuals in all cultures go through the same order or sequences of gross stages of development, though varying in rate and terminal point of development (Kohlberg, 1971, p.176)

While it is unclear which 30 basic moral categories Kohlberg referred to, he listed 26 aspects of moral judgment or valuing across three dimensions of modes, elements and issues or institutions in the same paper he made the claim; each stage in his model encompasses these three dimensions. Kohlberg claimed that any of these aspects can be found in all cultures. Table 10 presents these 26 aspects.
Table 10. Aspects of Moral Judgment*

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Aspects</th>
</tr>
</thead>
</table>
| Modes of judgment of obligation and value | Judgment of right  
Judgment of having a right  
Judgment of duty and obligation  
Judgments of responsibility – conceptions of consequences of action or of the demands or opinions of others one should consider over and above strict duties or strict regard for the rights of others  
Judgment of praise or blame  
Judgments of punishability and reward  
Justification and explanation  
Judgments of non-moral value or goodness |
| Elements of obligation and value | Prudence – consequences desirable or undesirable to the self  
Social welfare – consequences desirable to others  
Love  
Respect  
Justice as liberty  
Justice as equality  
Justice as reciprocity and contract |
| Issues or institutions | Social norms  
Personal conscience  
Roles and issues of affection  
Roles and issues of authority and democracy, of division of labour between roles relative to social control  
Civil liberties – rights to liberty and equality to persons as human beings, as citizens, or as member of groups  
Justice of actions apart from fixed rights – reciprocity, contract, trust, and equity in the actions or reactions of one person  
Punitive justice  
Life  
Property  
Truth  
Sex |

* Adapted from Kohlberg (1971, p.166)
Despite Kohlberg’s claims, Locke (2011) claimed that in his own research, Kohlberg had been unsuccessful in demonstrating this claim. On the contrary, Locke claimed that Kohlberg’s own research had inevitably provided evidence to support cultural relativism, rather than to support his claim that this set of moral categories are universal and culturally independent. Thus, while Kohlberg himself did not support the notion that the stages of moral reasoning could vary across cultures, this claim remains somewhat contested.

In support of the above position, various researchers who have attempted to verify Kohlberg’s claims of cultural invariance of the moral reasoning stages have produced mixed results. For example, Snarey (1985) conducted a meta-analysis of 45 cross-sectional and longitudinal studies of moral development across 27 countries deemed to represent different cultures. Results indicated that, while Kohlberg’s interviews were deemed culturally fair as they were adjusted to suit each cultural context, Kohlberg’s stages favoured populations from urban societies over small societies such as village tribes, because participants from the latter groups never attained the post-conventional levels. Other researchers, such as Baek (2002), who have focused on exploring potential issues of culture bias in Kohlberg’s Moral Judgement Interview, have reached similar conclusions. In studying how Korean children responded to moral issues, Baek found that their responses could not be fully matched to the scoring protocols used in Kohlberg’s Moral Judgment Interview. Baek opined that Kohlberg was unaware of certain moral reasoning dimensions attributable to culture and unique to Korean children, and that cultural background should be considered whenever interpreting children’s moral reasoning scores on this instrument.
Of direct relevance to the present study context, Fang, Fang, Keller, Edelstein, Kehle and Bray (2003) studied 350 Mainland Chinese (Beijing) students with an equal number of boys and girls aged between 9 to 15 years old. Based on their results, Fang et al. concluded that Chinese children’s moral decisions placed more of an emphasis on respect for authority and altruism than those of the subjects in Kohlberg’s original studies. Despite these results, Fang et al. argued that ethnicity, in this case Chinese, did not appear fundamentally to impact the stage invariant theory posited by Kohlberg.

Thus, at this point, empirical findings with respect to the cultural invariance of Kohlberg’s stages appear to provide some support for the notion that the model is insensitive to subtle cultural differences. It is noteworthy how Kohlberg himself responded to his critics on his claim as cited in Edwards (1986, p.427):

> We do not believe that the comparison of one culture to another in terms of moral development is a theoretically useful strategy for the growth of scientific knowledge. It is difficult to understand what a valid concept of 'comparative moral worth of culture' might be, but in any case such a concept could not be established on the basis of a comparison of means, on our moral judgment assessment scale. There is no direct way in which group averages can be translated into statements of the relative moral worth of groups (p. 113).

This statement makes clear that Kohlberg himself was not firm in his opinion that his model could be translated meaningfully across cultures. As such, this area is one in which further research is clearly needed.
1.4. GENDER AND MORAL REASONING

As noted previously, amongst the various critics of Kohlberg’s stage theory who have published in the field, one of the first was Gilligan, Kohlberg’s research assistant, who opined that Kohlberg’s theory is gender-biased. Gilligan’s point of contention was grounded primarily in the fact that Kohlberg’s research was based on an all-male sample, presented with scenarios that included only male protagonists, which made it difficult for females to relate to. Hence, Gilligan posed that Kohlberg’s developmental stages reflect a male definition of morality, which will be based on abstract principles of law and justice, rather than a more feminine perspective, which will be based on principles of compassion and care.

As noted previously, despite the face validity of Gilligan’s claims, empirical research has generally not supported her views. For example, Nunner-Winkler (1990) found no qualitative distinction between the kinds of responses made by males and females (aged 14 to 22) towards two moral dilemmas. Galotti (1988) also found little evidence to support differences in how women thought about moral dilemmas as compared to men. Similarly, in concluding that gender differences in Kohlberg’s theory were not significant, Donenberg and Hoffman (1988) found from an analysis of variance that while girls emphasised morality based on care more than justice, boys emphasised morality based on justice and care equally, contrary to Gilligan’s assertions. In Singapore, Koh (2012) found from her study involving 183 Singapore students aged 13 to 17 years that the “development of moral reasoning follows a similar trend in males as in females” (p.94). Based on these results, assertions of gender differences in moral reasoning
have not met with strong empirical support thus far. Despite this, it is clear from the opponents of the Kohlberg instrument that gender differences should be explored in the development of any instrument designed to assess moral reasoning levels.

1.5. THE PRESENT STUDY CONTEXT

Various scholars in the field of moral reasoning have noted the importance of efforts within education to develop students’ moral reasoning abilities, particularly during the adolescent years. For example, Vera-Estaya, Dooley and Beauchamp (2015) asserted that moral reasoning is an “important skill during adolescence because it guides social decisions between right and wrong” (p.17). Others have noted that moral reasoning is a “basic component of moral education, character education and education for democracy” (Zarinpoush, Cooper, & Moylan, 2000, p.397), and that “knowledge of moral reasoning is one important component for building educational programs that can help reduce future unwanted behaviour” (Backman & Gardelli, 2015, p.227).

Given that one of the desired outcomes of education specified by the Singapore Ministry of Education (MOE) since 1997 has been that individuals who go through the Singapore education system become confident people who are discerning in judgment, and possess a strong sense of right and wrong, moral reasoning has now become a core aspect of Singapore’s current Character and Citizenship Education (CCE) syllabus.
As the most influential theory of moral development, Kohlberg’s theory is espoused specifically in the CCE syllabus as a pedagogical approach that teachers can use in supporting this curriculum. In addition to recommending the use of the Kohlberg framework to teachers, the CCE syllabus also stipulates baseline stages within the Kohlberg framework that students are expected to attain at different ages:

1. Primary 1-2 students should be able to reason up to pre-conventional stage 2.
2. Primary 3-5 students should be able to reason up to conventional stage 3.
3. Primary 6 to Secondary 2 students should be able to reason up to conventional stage 4.
4. Secondary 3 to 5 students should be able to reason up to post-conventional levels, which includes stages 5 and 6 (i.e., students at this level are expected to show principled moral reasoning).

The MOE suggests that teachers at the secondary level assist students in reaching these levels by discussing moral dilemmas with students, and modelling how informed moral decisions should be made in the context of these dilemmas. This approach is supported by Tan and Chew (2004), who similarly detailed how such pedagogies can be used:

Through discussions and the resolving of values conflict and priorities, pupils are meant to move from initially self-regarding motives for values choice towards more altruistic, other regarding reasoning, in line with
recommendations of cognitive developmental psychologists, such as Piaget and Kohlberg (p.604).

Thus, in the Singapore context, teachers are increasingly being tasked with responsibility for developing students’ moral reasoning skills, based on Kohlberg’s theory. Despite this initiative, at present, there is no standardised approach to assessing progression towards the higher stages within the model that is suitable for use with large groups of secondary level students. Although two widely used instruments are available to assess moral reasoning development based on Kohlberg’s model (Kohlberg’s original Moral Judgement Interview and the Defining Issues Test – 1 and 2), neither provide a practical means by which large numbers of students can be assessed. Thus, while teachers are now being encouraged to facilitate students’ growth within this area, there is at present no practical, systematic instrument available to them to determine how well students are actually progressing on the basis of their efforts.

1.6. OVERVIEW OF THE PRESENT STUDY

Given the importance of moral reasoning within the Singapore Ministry of Education CCE curriculum, and its basis in Kohlberg’s (1984) Theory of Moral Development, the aim of the present study was to develop an instrument to measure moral reasoning levels on a broad-scale basis within Singapore secondary classrooms. The instrument developed and validated in the present study was named the Moral Reasoning Questionnaire, or MRQ. The MRQ is based on Kohlberg’s (1984) theoretical model, with items presenting obligatory moral
dilemmas that integrate prosocial, anti-social and social pressure elements to reflect situations familiar to Singapore secondary level students. The Chinese population in Singapore is the largest and hence, efforts were made to research Asian moral reasoning theories in addition to Kohlberg’s. Nonetheless, there is currently a lack of Asian grounded moral reasoning frameworks.

Chapter 2 of this thesis presents a review of existing instruments to assess moral reasoning in children, adolescents, and adults. This review highlights the limitations of each instrument for use on a large-scale basis within secondary schools, providing the rationale for development of the MRQ. Chapters 3 and 4 describe the procedures used in developing and validating the MRQ within the Singapore secondary school contexts, and the results of the preliminary validation study conducted. Chapter 5 then discusses the practical implications of, and directions for future research suggested by, the results of the study.
Chapter 2. Assessing Moral Reasoning

This chapter discusses how moral reasoning has historically been assessed. According to Palmer (2011), measures of moral reasoning can be categorised into two main types: (i) production measures, which require a participant to construct a response to justify or support a decision, and (ii) recognition measures, which require a participant to recognise and select a response to justify or support a decision (which can be a rate, rank or forced-choice item that closely matches what the participant agrees or disagrees with).

These two types of measures are similar to cognitive tests that require an examinee to respond to selected or constructed response items. Constructed response items, in general, provide more information on an examinee’s understanding and mastery of given concepts. However, more time is required to complete a test with constructed response items, and scoring is more complex, with markers’ subjectivities producing bands of error in the assignment of scores to performances. While marker subjectivities can be minimised, more time is generally required for scoring constructed response items.

Selected response items often do not provide as much information as constructed response items, but are generally more practical if an instrument is to be used in large-scale assessments. Given these practical advantages, a selected response format was chosen in the development of the instrument within the present research. In this development process, care was taken to ensure that the potential disadvantages of the selected response approach were minimised to the extent
possible. For instance, in areas as complex as moral reasoning, it is essential that the options used in selected response items be developed in consultation with an expert panel, and trialled with a small sample of potential respondents, before it is used in a larger sample. Both of these steps were taken in validating the instrument developed here.

The remainder of this chapter focuses upon providing a critical review of existing instruments to assess students’ moral reasoning levels. Both advantages and disadvantages of each instrument are highlighted, as these ultimately formed the rationale for the development of the new instrument in the present study.

2.1. MORAL JUDGMENT INTERVIEW

Originally, Kohlberg presented the Moral Judgement Interview (MJI) in his 1958 dissertation, which had been developed as an operationalisation of his Theory of Moral Development. The MJI is a production measure of moral reasoning. Its administration involves interviewing students in a semi-structured format, with each interview lasting between 30 to 90 minutes (depending on how the interviewee responds and how long the interviewer persists with specific questions in the interview - see Gibbs et al., 1982). In the interview, the interviewer uses nine hypothetical moral dilemmas to determine which stage of moral reasoning the interviewee is applying when considering his/her response to each dilemma.
The moral dilemmas are fictional short stories that describe situations in which a person has to make a moral decision. The respondent is then asked a series of nine to twelve standardised and prescribed open-ended questions to elicit what they think the right course of action is, and, most importantly, why. Respondents are not asked what they would do or how they would act in that moral dilemma. The form and structure of these replies is then scored, rather than the content. As an example, the Heinz dilemma was one of the moral problems that Kohlberg used in his very first studies:

A woman was near death from a special kind of cancer. There was one drug that the doctors thought might save her. It was a form of radium that a druggist in the same town had recently discovered. The drug was expensive to make, but the druggist was charging ten times what the drug cost him to produce. He paid $200 for the radium and charged $2,000 for a small dose of the drug. The sick woman's husband, Heinz, went to everyone he knew to borrow the money, but he could only get together about $1,000 which is half of what it cost. He told the druggist that his wife was dying and asked him to sell it cheaper or let him pay later. But the druggist said: “No, I discovered the drug and I'm going to make money from it.” So Heinz got desperate and broke into the man's store to steal the drug for his wife. Should Heinz have broken into the laboratory to steal the drug for his wife? Why or why not?

The MJI has three standard forms (Forms A, B and C), and stage scoring for each form is conducted on the basis of the Standard Issue Scoring Manual. It should be noted that in the scoring of the responses, what the participant thinks that Heinz should do is not important. It is the justification the participant offers for his/her
choice that is important. In constructing the MJI, Kohlberg was not concerned with the interviewees’ judgments *per se* - his aim was to map their moral *reasoning* to the different stages within his model. Kohlberg assumed that the reasoning exhibited by respondents formed the basis for their judgments. Examples of possible arguments that correspond each to the six stages are:

- **Stage one (obedience):** Heinz should not steal, because he would then be imprisoned, which would make him a bad person.
- **Stage two (self-interest):** Heinz should steal, because, even though he might go to prison, saving his wife would make him happier.
- **Stage three (conformity):** Heinz should steal, because this would make him a good husband - he would be meeting his wife’s expectations.
- **Stage four (law-and-order):** Heinz should not steal, because it is against the law.
- **Stage five (human rights):** Heinz should steal, because everyone has a right to medications that are life-saving, even if this is against the law.
- **Stage six (universal human ethics):** Heinz should steal, because the right to live is more fundamental to society than intellectual property rights.

In scoring the MJI, a respondent can be awarded two indices. The *global stage score* is given on a 9-point scale (five stages of moral development plus four
transitional stages). The weighted average, the moral maturity score, can also be
given, which ranges from 100 (fully stage one) to 500 (fully stage five).

The scoring process for the MJI is complex and time-consuming to complete,
because it relies on scorers’ judgements of the stage that best fits a given
interviewee’s response to each of the dilemmas. In fact, Miller (2007) wrote that
the MJI scoring might be the most complex scoring system in the field of
psychology. Furthermore, respondents who participate in ‘poorly probed’
interviews often need to be awarded ‘guess’ scores by scorers, because their
reasoning can be unclear based on the responses received (Colby et al., 1983).
Various scholars have questioned the reliability of the MJI based on these post-
interview scoring practices. Cortese (1984), for example, highlighted some of the
potential problems caused by ‘guess’ scores, which then factor into the overall
MJI indices. ‘Guess’ scores can be related to how the interviewer conducts the
MJI, and also, a lack of understanding or misunderstanding of the stage
structures within the instrument, especially of the nine or thirteen category stage
scoring. Cortese (1984) further added even trained interviewers might disagree
on the number of ‘whys’ an interviewee should be asked for each dilemma (i.e.,
the extent to which responses should be ‘probed’), and that an interviewer’s
choice of words would inevitably influence the interviewee’s responses.

Despite the criticisms of the MJI scoring system, a small number of empirical
evaluations have provided mixed support for the reliability and validity of the
MJI. Colby et al. (1983), for example, reported high test-retest reliabilities
of .98, .96 and .92 for the MJI Forms A, B and C, based on a one-month test-retest
interval. In the same study, factor analyses and Cronbach’s alphas also supported the unidimensionality and high internal consistency of the MJI (\(\alpha_{Form\ A} = .92\), \(\alpha_{Form\ B} = .96\) and \(\alpha_{Form\ C} = .90\)). Inter-rater reliability, however, was found to be less robust, ranging from .53 to 1.00 across the three forms, depending on the level of specificity of the judgements being made. Taking Form A as an example, inter-rater agreement percentages ranged from 88% to 100% for agreement within a third of a stage, from 75% to 88% for complete agreement based on the nine category scale (stage 1, 1/2, 2, 2/3, 3, 3/4, 4, 4/5, 5), and from 53% to 63% for complete agreement using the most granular 13 category scale (1, 1/2, 2/1, 2, 2/3, 3/2, 3, 3/4, 4/3, 4, 4/5, 5/4, 5).

Despite the mixed positive evidence cited by researchers to support the MJI, the potential impact of interviewer subjectivity upon the scoring remains a significant stumbling block to the use of this instrument. It is apparent that the validity and reliability of the MJI hinges not only on the interviewee’s ability to articulate his or her reasoning, but also, on the skill of the interviewer in eliciting interpretable responses. In a situation in which an interviewee is less articulate than others, or in situations where interviewers have a tendency to under- or over-interpret conversations, the validity of the MJI scores would ultimately be compromised. Thus, while the MJI has been used extensively and has been reported to exhibit high levels of internal consistency and test-retest reliability (Colby et al., 1983), alternatives to mitigate practicality issues such as the time required to do the test, reliance on highly trained interviewers, the complex scoring and coding system, and some of the inter-rater and validity issues associated with the test, are needed.
2.2. **Defining Issues Test**

According to Miller (2007), there are two alternatives to Kohlberg’s MJI: (i) the *Defining Issues Test* (DIT); and (ii) the *Sociomoral Reflection Measure* (SRM). The latter measure is considered in more detail in a later section. Both the DIT and the SRM have been demonstrated to have favourable psychometric properties, though the DIT has been reported to exhibit a lower correlation with MJI scores than the SRM (Palmer, 2011). Miller (2007) distinguished between the DIT and the SRM in the following way:

In the SRM, the dilemmas are presented in writing and the participant responds in writing, whereas in the DIT the participant ranks various issues with regard to how important each is in deciding how the dilemma ought to be resolved. Both measures are less time-consuming than is the standard Kohlberg approach, and each is easier to administer and to score (p.284).

In contrast to the MJI, the DIT is a recognition measure of moral reasoning, originally developed by Rest (1979). The DIT is generally considered the primary alternative to Kohlberg’s MJI for assessing stages in the original Kohlberg model. Though Rest (1979) stated that he did not intend the DIT to be considered an optimal measure of moral reasoning, and encouraged further explorations of its properties in the original validation of the instrument, the DIT is now a prominent alternative to the MJI (Gibbs et al., 1982). It is likely that the popularity of this instrument stems in part from the fact that it is less time-consuming and expensive to administer than the MJI.
There are six moral dilemmas in the original DIT and three in an abbreviated DIT (Elm & Weber, 1994). Respondents to the DIT read a series of moral dilemmas and, against each dilemma, rate 12 stage-related factors that could be considered in responding to the dilemma in terms of their importance, on a five-point scale (from of great to no importance). Respondents then select and rank four of these 12 items as most to least important (see Figure 2 for an example).

According to Rest (1979), the DIT assumes that people at different developmental levels in their moral reasoning will respond to each moral dilemma differently. Taking Heinz’s dilemma as an example (a dilemma that used both in the MJI and in the DIT), some may subscribe more to upholding social order and maintaining the laws of a community; others may consider a Heinz’s love for his wife as paramount; and others may consider Heinz’s self-preservation as critical.

In rating and ranking the stage-related considerations following the moral dilemmas, the DIT assumes that a respondent has used a specific stage or at most two adjacent stages of moral reasoning in making his or her response, noting that a subsequent stage is a reconstruction or transformation of the previous. Rest (1979) designed the DIT as a developmental measure of moral judgment by a two-stage process of preference and recognition. Thus, the DIT is not reliant on expressive verbal skills, unlike the MJI (although it does rely on the respondent’s ability to read the dilemmas and ranking options).
HEINZ AND THE DRUG

In Europe a woman was near death from a special kind of cancer. There was one drug that doctors thought might save her. It was a form of radium that a druggist in the same town had recently discovered. The drug was expensive to make, but the druggist was charging ten times what the drug cost to make. He paid $200 for the radium and charged $2,000 for a small dose of the drug. The sick woman's husband, Heinz, went to everyone he knew to borrow the money, but he could only get together about $1,000, which is half of what it cost. He told the druggist that his wife was dying, and asked him to sell it cheaper or let him pay later. But the druggist said, "No, I discovered the drug and I'm going to make money from it." So Heinz got desperate and began to think about breaking into the man's store to steal the drug for his wife.

Should Heinz steal the drug? (Check one)

______ Should steal it   ______ Can't decide   ______ Should not steal it

**IMPORTANCE:**

<table>
<thead>
<tr>
<th>Great</th>
<th>Much</th>
<th>Some</th>
<th>Little</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From the list of questions above, select the four most important:

Most important _______ Second most important _______

Third most important _______ Fourth most important _______

A respondent’s level of moral (reasoning) development is represented by the P-score, an index based on the respondent’s ranking of post-conventional items regardless of how he or she ranks other stage-related items. For example, after reading a moral dilemma, a respondent ranks the four most important stage-related items out of the 12 given items. The item ranked most important is given a score of four and the one ranked least important is scored one. Scores are then totalled across all six moral dilemmas and divided by .60 to give a percentage score given that the maximum scored attainable is 60. Scores related to ranking of post-conventional items are then summed to generate the P-score, which is then used to represent the respondent’s level of moral reasoning development.

While there were concerns that score inflation or deflation might occur when respondents fake good (high) or bad (low), Rest (1979) cited various studies which suggested that faking would not impact DIT scores significantly. Further, the DIT includes meaningless statements that appear philosophical as ‘foil’ items. Score interpretation for respondents who consistently select these statements would be qualified or not attempted, because these selections would be deemed to suggest a respondent attempting to inflate his/her score artificially.

The validity and reliability of the DIT has been well-established (Elm & Weber, 1994). For example, works by Rest, Thoma, Narvaez and Bebeau (1997) suggest adequate to good internal reliability of the DIT (P-score), with Cronbach’s $\alpha$ in the range of .76 for a 1979 composite sample ($N=994$) to .78 for a 1994 composite sample ($N=932$). In reviewing the empirical evidence relating to the validity of the DIT, Rest, Thoma and Edwards (1997) concluded that the support for the
instrument was strong, based on evidence garnered across seven construct validity criteria:

(i) differentiation of P-scores between age-education groups;
(ii) longitudinal gains (in P-scores);
(iii) correlation with cognitive capacity measures;
(iv) sensitivity to moral education interventions;
(v) correlation with behaviour and professional decision making;
(vi) relations with political choice and attitude (e.g., by comparing DIT scores with other validated instruments established to measure political attitude e.g., Rokeach Values Survey; and
(vii) “fakeability” studies.

Rest (1979) cited numerous studies that had provided empirical evidence to support the validity of the DIT as a moral development measure, including:

(i) Results of a study of the internal structure of the DIT in a sample of 160 subjects, and later cross-validated on another sample of 1080 subjects, which indicated that the DIT scale values were ordered according to Kohlberg’s theoretical framework;
(ii) A cross-sectional study of different groupings by age-education (junior highs, senior highs, college and graduate students – a total of over 4500 students across the entire United States), which showed that age-education accounted for 38% of variance of moral judgment scores;
(iii) A series of longitudinal studies, which indicated that individuals generally achieved higher moral judgment P-scores as they aged;
(iv) A test-retest study on the same subjects in 1972, 1974 and 1976, which indicated that mean P-scores increased from 33 to 44 over time; and

(v) Correlations with IQ, aptitude and achievements scores such as the SAT tests in the .20 to .50 range, which suggested that the DIT was neither a substitute for cognitive tests nor a cognitive test that could measure moral development.

Despite the favourable empirical evidence cited above, two significant criticisms of the DIT have appeared regularly within the literature on moral reasoning measurement: (i) the issue of using a quantitative measure to describe a qualitative theoretical framework, and (ii) using the P-score as a reflection of an individual’s moral development.

In responding to the qualitative-quantitative criticism, Rest et al. (1997) asserted that the DIT is developmental and not evaluative. Rest (1979) stated that the construction and validation of the DIT served to provide information for the exploration of moral judgment theories. It was noted that a DIT score was not intended to place an individual definitively in a particular stage, and that individuals do not consistently belong to one unique stage. Rather, researchers designed the DIT to measure the “extent of and under what conditions does a person manifest particular stages of thinking” (Rest et al., 1997, p.499).

Rest et al. acknowledged, with respect to using the P-score as a reflection of an individual’s moral development, that there was a loss of information in not considering how an individual ranks non post-conventional items (Rest, Thoma, Narvaez, & Bebeau, 1997; Thoma & Dong, 2014). To overcome this, a new index,
the N2-score, was developed and published in 1997 (Rest et al., 1997). The N2-score is an adjusted P-score based on a respondent’s ability to discriminate between post-conventional related items and lower stage items. This score increases in a positive direction if a respondent is able to discriminate between higher and lower stage related items. As the N2-score is anchored on the P-score, it correlates well in the region of .90 with the P-score.

In his study of another one hundred studies and in constructing the DIT, Rest (1979) concluded that moral judgment is developmental, though the age range for each stage or level of development may vary for different theoretical frameworks. He also concluded that a major source of variation other than age is social experience. Indeed, social experience and hence familiarity plays a role in how subjects respond to items. The dilemmas in the DIT seem non-familiar and dated as daily experiences to students at the secondary level or equivalent.

In response to problems related to the DIT such as the P-score that does not consider non post-conventional ranking, dated dilemmas, potential respondent “fakeability”, and group administrability, a revised version, the Defining Issues Test – 2 (DIT2) was published in 1999. The DIT2 (see example item in Figure 3) is a selected response recognition measure which provides quantitative scores based on the test-taker’s responses to five hypothetical moral dilemmas; it is purported to measure moral reasoning based on Kohlberg’s original model and is parallel to the DIT in construction (Rest, Narvaez, Thoma & Bebeau, 1999b).
The small village in northern India has experienced shortages of food before, but this year’s famine is worse than ever. Some families are even trying to feed themselves by making soup from tree bark. Mustaq Singh’s family is near starvation. He has heard that a rich man in his village has supplies of food stored away and is hoarding food while its price goes higher so that he can sell the food later at a huge profit. Mustaq is desperate and thinks about stealing some food from the rich man’s warehouse. The small amount of food that he needs for his family probably wouldn’t even be missed.

What should Mustaq Singh do? Do you favor the action of taking the food? (Check one)

<table>
<thead>
<tr>
<th>Strongly Favor</th>
<th>Favor</th>
<th>Slightly Favor</th>
<th>Neutral</th>
<th>Slightly Disfavor</th>
<th>Disfavor</th>
<th>Strongly Disfavor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Rate the following issues in terms of importance (1 = great, 2 = much, 3 = some, 4 = little, 5 = no). Please put a number from 1 to 5 alongside every item.

1. ☐ Is Mustaq Singh courageous enough to risk getting caught for stealing?
2. ☐ Isn’t it only natural for a loving father to care so much for his family that he would steal?
3. ☐ Shouldn’t the community’s laws be upheld?
4. ☐ Does Mustaq Singh know a good recipe for preparing soup from tree bark?
5. ☐ Does the rich man have any legal right to store food when other people are starving?
6. ☐ Is the motive of Mustaq Singh to steal for himself or to steal for his family?
7. ☐ What values are going to be the basis for social cooperation?
8. ☐ Is the epitome of eating reconcilable with the culpability of stealing?
9. ☐ Does the rich man deserve to be robbed for being so greedy?
10. ☐ Isn’t private property an institution to enable the rich to exploit the poor?
11. ☐ Would stealing bring about more total good for everybody concerned or not?
12. ☐ Are laws getting in the way of the most basic claim of any member of a society?

Which of these 12 issues is the 1st most important? (Write the number of the item) __________________

Which of these 12 issues is the 2nd most important? __________________

Which of these 12 issues is the 3rd most important? __________________

Which of these 12 issues is the 4th most important? __________________

Figure 3. Example of a DIT2 item*

Similar in nature to the DIT but shorter and with clearer instructions, each of the five hypothetical moral dilemmas in the DIT2 is followed by 12 issues that could be considered in resolving the dilemma. Participants are asked to indicate a decision in each dilemma based on a seven-point scale ranging from strongly favour to strongly disfavour; they then indicate which of the listed issues are most important to their decision using a five-point scale ranging from great to no importance. Responses are scored to quantify, based on the Schema Theory, which of the following three moral schemas that the individual had used in coming to his/her decision: (i) personal interest schema (i.e., thinking about what will benefit him/her or making others like him/her); (ii) maintaining norms schema (i.e., thinking about which decision would maintain law and social order); or (iii) post-conventional schema (i.e., thinking about human rights and other higher moral principles). The index characterising the moral development of an individual is the N2-score, an improved index over the P-score from the DIT.

Owing to its comparatively limited history, the DIT2 has a less extensive empirical evidence base to support it than the original DIT. Rest and colleagues (1999b) have, however, validated the DIT2 based on similar criteria to the seven used to validate the original DIT. These validation studies on the DIT2 have indicated:

(i) Significant age and educational differences across the DIT2 N2-scores among ninth graders (n=47), high school graduates (n=35), college seniors (n=65), and students in graduate and professional schools (n=53) as with the DIT P-scores (see Table 11).
(ii) The DIT2, as with the DIT, was related to views on public policy issues. Studies supported a slightly higher correlation between the DIT2 and the Attitudes toward Human Rights Inventory (ATHRI) \((r=.50)\), relative to that with the DIT \((r=.48)\); the ATHRI is an instrument where respondents agree or disagree on a five-point Likert scale with statements related to public policy issues drawn from American Constitution’s Bill of Rights including abortion, euthanasia, homosexual rights, due process rights of the accused, free speech, women’s roles, and the role of religion in public schools. Studies also supported a higher partial correlation of the DIT2 with ATHRI \((r=.51)\) relative to that with the DIT \((r=.40)\). The correlations were partial as religious fundamentalism and political conservatism were controlled in view that these two factors correlated strongly with the DIT in previous studies by Rest. Religious fundamentalism was based on the Inventory of Religious Belief, a 15-item measure that uses a five-point Likert scale to distinguish those who believe and reject the literalness of Christian tenets, and political conservatism was measured on the political conservatism scale where respondents identified their political identity based on a five-point scale.

(iii) Adequate internal reliability of the DIT2. Cronbach’s alpha for the DIT2 N2-score was .81 \((n=192)\) while that for the DIT P-score was .76 \((n=154)\). Further, the DIT2 N2-score had a high correlation with the DIT P-score \((r=.71)\) \((n=154)\), and an even higher correlation when N2-scores for the DIT were computed and used \((r=.79)\) \((n=178)\).
Table 11. Means and Standard Deviations of DIT P-score and DIT2 N2-scores

<table>
<thead>
<tr>
<th>Education level</th>
<th>P-score (n=154)</th>
<th>N2-score (n=192)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Ninth grade</td>
<td>23.0</td>
<td>10.0</td>
</tr>
<tr>
<td>College freshmen</td>
<td>28.7</td>
<td>11.5</td>
</tr>
<tr>
<td>College seniors</td>
<td>33.7</td>
<td>14.1</td>
</tr>
<tr>
<td>Graduate/professional school</td>
<td>53.9</td>
<td>13.1</td>
</tr>
</tbody>
</table>

Adapted from Rest et al. (1999b, p.7)

In a more recent review of the DIT/DIT2, Thoma and Dong (2014) noted that numerous specific criteria had been applied to validate the DIT/DIT2 across studies: (i) the extent to which it differentiated across various age/education groups; (ii) patterns of longitudinal gains made on the DIT/DIT2; (iii) correlations between DIT/DIT2 scores and those on cognitive ability measures; (iv) the ability of the DIT/DIT2 to detect the effects of moral education interventions; (v) correlations between the DIT/DIT2 scores and decisions made with respect to students’ behaviour; and (vi) correlations between the DIT/DIT2 scores and other measures such as political choice and attitude. Thoma and Dong, having reviewed the evidence available on the validity of the DIT/DIT2 based on these criteria, concluded that the empirical evidence had largely supported the DIT/DIT2 as a valid measure of moral reasoning. Thus, in comparison to the MJII, the DIT/DIT2 is considered by many to have a stronger evidence base.

Despite evidence to support the validity of the DIT and later the DIT2, criticisms remain. An early critique was proffered by Kay (1981), who asserted that Rest and colleagues used correlational designs confounded by extraneous variables in their cross-sectional and longitudinal studies, and the failure to isolate study variables was a severe limitation of the studies. In cross-sectional studies, P-
scores measured could be inaccurate due to the covariances of other significant variables such as social economic status and moral training. In longitudinal studies, P-score gains could well be impacted by age or educational levels alone, or by any other variables such as moral training that have impact on either or both of the two. This assertion undermines the notion of the DIT as an instrument that measures moral development exclusively.

A more recent critique of the DIT and DIT2 was by Curzer, Sattler, DuPree and Smith-Genthōs (2014), in their attempt to develop an alternative to the DIT called the Sphere-Specific Moral Reasoning and Theory Survey. Curzer et al. (2014) asserted that anchoring any instrument on a single theory is flawed, because: (i) moral development is uneven; (ii) there is no one universal moral theory that is agreed upon by all experts; and (iii) respondents are inconsistent and apply varying moral theories in solving dilemmas. This assertion drew a strong response and rebuttal from DIT researchers. Thoma, Bebeau and Narvaez (2016) stated categorically that Curzer and colleagues did not “consult existing empirical evidence, misunderstand the model and method associated with the Defining Issues Test, and thereby reach conclusions that are unwarranted, incomplete, and ultimately indefensible” (p. 241). Thoma et al. (2016) stated that DIT researchers had all along acknowledged that the DIT is not a comprehensive measure of moral development but only provides a picture of developmental features of moral judgment. While DIT researchers agreed that there are alternative moral theories, they suggested that it was reasonable to anchor the DIT and DIT2 on Kohlberg’s model and the Schema Theory given that data from sample after sample showed moral development in the assumed models.
2.3 Ethical Reasoning Inventory

The Ethical Reasoning Inventory (ERI), developed and validated by Page and Bode (1980) is similar in content and foci to the DIT and DIT2, but is simpler to administer and score. The ERI was developed partly in response to the poor internal consistency data reported for the original DIT, in comparison to those reported previously for the MJI (e.g., Cronbach’s alpha at the dilemma level for the DIT .65, compared to .89 for the MJI). The ERI requires participants to respond to six dilemmas similar to those used by Kohlberg in the MJI. For each dilemma, participants first select one of two ‘action’ options. Based on their selected response, participants turn to the relevant page and select, out of six options corresponding to Kohlberg’s six stages, an option that best represents their reasoning for choosing the action they did. Thus, participants need to select two options for each dilemma presented.

The ERI has not been validated on the basis of the seven criteria proposed by Rest and colleagues. Comparisons between different measures of moral reasoning have, however, suggested that the ERI is a more reliable instrument than both the MJI and the DIT. For example, in one study by Page and Bode (1980) in which the MJI, DIT and ERI were all administered in a sample of college freshmen and sophomores (N=92), the correlation between scores of the MJI and ERI was higher (r=.54) than that between the MJI and DIT (r=.50). Further, the coefficient of stability of the ERI under test-retest conditions of college students (N=51) was .69 with an interval of 10 days; this was slightly higher than that of the DIT at .65 (N=47) with an interval of 18 days. Thomas (1986) stated that a higher coefficient
of stability points to a more consistent assessment technique over time. Hence, the ERI in this instance can be considered more consistent than the DIT. It should be noted, however, that the sample of college students used for comparison was not the same.

While Page and Bode (1980) used Pearson correlation coefficients to demonstrate superior consistency of the ERI, more empirical validation studies could have been conducted. The dilemmas used were similar to those used in Kohlberg’s MJI, but there was no mention that the ‘action’ options and the latter six options were deemed content-appropriate by experts. As with the DIT, there was also no discussion establishing the internal factorial structure of the ERI using classical test theory methods such as factor analysis.

Nevertheless, Bode and Page (1979) stated that the ERI possesses reliability, construct validity and can be used with subjects aged as young as 14 based on their definition of validity. In addition, their “fakeability” studies on 174 college students suggested that respondents were unable to fake upwards though they could fake downwards significantly (Page and Bode, 1979). While the ERI is more easily group-administrable compared with the MJI, Rest et al. (1997) stated that higher alpha values may not mean an instrument is better than another, as opposed to Page and Bodes’ (1980) claim that the ERI has better reliability coefficient of internal consistency. Given the lack of literature on the ERI and the existing literature on the DIT and DIT2, it cannot be concluded that the ERI is a superior instrument to the DIT for measuring moral reasoning development.
2.4. Sociomoral Measures

More recently developed and also widely used instruments to measure moral reasoning based on Kohlberg’s stages of moral development include the Sociomoral Reflection Measure (SRM) developed by Gibbs et al. (1982) and its derivatives (i.e., the Sociomoral Reflection Objective Measure, or SROM, developed by Gibbs, Arnold, Morgan, Schwartz, Gavaghan, & Tappan, 1984; the SROM-Short Form, or SROM-SF, developed by Basinger & Gibbs, 1987; the Sociomoral Reflection Measure – Short Form, or SRM-SF, developed by Gibbs, Basinger and Fuller, 1992; and the most recent Sociomoral Reflection Measure – Short Form Objective, or SRM-SFO, developed by Brugman, Basinger & Gibbs, 2007).

The initial SRM was an attempt to make the MJI more group administrable in light of criticisms of the MJI which requires a substantial investment of time, effort, and cost for its effective use (e.g., interviewers had to attend a five to ten-day workshop just to learn about the MJI scoring structure) (Gibbs et al., 1982). In addition, Gibbs et al. (1982) cited evidence that the primary alternative to the MJI, the DIT, did not correlate adequately with the MJI when chronological age was partialed out. Hence, Gibbs and colleagues sought to develop the SRM, an MJI equivalent that is more feasible in terms of administration.

Similar to the MJI, the SRM is a production measure and assesses justificatory moral judgment that lasts between 45 to over 60 minutes; this is about 30 minutes less than the MJI requires. In the SRM, respondents reflect and express their thoughts on moral dilemmas similar to those used in the MJI. A key difference is that probe questions in the SRM were modified from the MJI so that these
questions may be more consistently efficacious, accommodate sufficient scorable responses and hence minimise “guess” scores, an issue the MJI faces; the modification was such that within every probing question, a respondent would be asked his/her moral reasoning and judgment. Using the Heinz moral dilemma as a stimulus, examples of the modified SRM probing questions are:

1. What if Heinz doesn't love his wife? Should Heinz: steal / not steal / can't decide (circle one)?
   1a. How important is it for a husband to steal to save his wife, even if he doesn't love her (whichever one you circled)?
   1b. WHY is that very important / important / not important (whichever one you circled)?

2. What if the person dying isn't Heinz's wife but instead is a friend (and the friend can get no one else to help)? Should Heinz: steal / not steal / can't decide (circle one)?
   2a. How important is it to do everything you can, even break the law, to save the life of a friend? very important / important / not important (circle one)?
   2b. WHY is that very important / important / not important (whichever one you circled)?

3a. What about for a stranger? How important is it to do everything you can, even break the law, to save the life of a stranger? very important / important / not important (circle one)?
   3b. WHY is that very important / important / not important (whichever one you circled)?
Scorers/raters have to undergo a training session before scoring the SRM. Nonetheless, the training can be conducted in a minimum of six hours which is relatively more manageable than that associated with the MJI training, which lasts a minimum of five days. Respondents will be scored based on their views on the probing questions; prescribed norms are tied to each probing question. Essentially, the SRM involves assessing respondents on eight sociomoral norms: (i) affiliation (marriage and friendship); (ii) life; (iii) law; (iv) legal justice; (v) conscience; (vi) family affiliation; (vii) contract; and (viii) property. For example, in the Heinz moral dilemma example above, justificatory moral judgment to questions 1b and 2b are related to the affiliation norm. Respondents’ mean stage level of responses to the eight sociomoral norms will be used to compute the two indices provided by the SRM: (1) the modal stage score and, (2) the Sociomoral Reflection Maturity Score (SRMS) that ranges from 100 to 400, analogous to the MJI’s Moral Maturity Score (MMS) though it is more differentiated psychometrically.

In validating the SRM, Gibbs et al. (1982) involved a sample of 107 subjects (59 female), aged from 12 to 22 (mean age = 15.5 years old) and studied four kinds of reliability: (1) inter-rater (2) test-retest (3) parallel form, and (4) internal consistency. Rater modal stage agreement ranged between 71 and 96 (excluding those involving Rater 10 as Rater 10, an undergraduate, was found to be the least skilled rater) and the SRMS inter-rater correlation ranged between .51 and .98 based on a sub-sample of 28. Noting that the lowest SRMS correlation of .51 was attributed to Rater 10, and most of the correlation coefficients that included this rater were low, it was concluded that the SRM had acceptable inter-rater
reliability. Further, test-retest (based on 52 of the 107 subjects) and parallel form (based on the remaining 55 subjects) correlation coefficients ranged from .70 to .90 indicating adequate reliability. Cronbach’s alphas were computed for the scores across different grades of all 107 subjects (grade seven, ten and college), and found to range between .58 and .75, with an overall alpha value of .85, indicating adequate internal consistency. In studying the construct validity of the SRM, Gibbs and colleagues found that the SRM covaried significantly with expected variables (age, grade and social economic status, or SES) and did not covary with sex. This provided evidence of the construct validity of the SRM and suggested that the instrument was not gender biased. While all these sources of evidence support the validity of the SRM (Colby et al., 1983), various derivatives were then developed to further mitigate against practical limitations associated with administration (Gibbs et al., 1984).

The SROM was developed as a recognition measure using forced-choice responses. It is a derivative of the SRM but requires about 20 minutes less to administer and less time to score (Basinger and Gibbs, 1987). In the SROM, participants have to technically rate and rank statements similar to the DIT and DIT2 after reading moral dilemmas also used in the MJI and SRM. Respondents have to complete 16 multiple-choice arrays (two arrays for each of the eight sociomoral norms similar to that used in the SRM). Using Heinz’s moral dilemma as an example, a stimulus array for the affiliation norm is:

1. What if Heinz's wife asks him to steal the drug for her?
   Should Heinz: steal/not steal/not sure (circle one)?
la. How important is it for a husband to do what his wife asks, to save her by stealing, even when he isn't sure whether that's the best thing to do? very important/important/not important (circle one)

lb. Let's say you had to give a reason WHY it is IMPORTANT for a husband to do that. What reason would you give? Is any of the following reasons close to the one you would give? (If a reason is too hard to understand, seems silly, or makes no sense, just circle "not close," or not sure).

   a. because it's his wife, and she told him to do it, so he should do what she says. close/not close/not sure (circle one)
   b. because he married her and if he didn't want to help her, why did he marry her in the first place? close/not close/not sure (circle one)
   c. because they may have formed together a deep mutual commitment. close/not close/not sure (circle one)
   d. because a good husband is expected to help his wife through sickness and health. close/not close/not sure (circle one)
   e. because he cannot recognize her without acceptance. close/not close/not sure (circle one)
   f. because he has accepted a responsibility as her husband. close/not close/not sure (circle one)

   Of all the above reasons, the reason which is the closest to the reason that you would give (or the least far off from the reason that you would give) is: a/b/c/d/e/f (circle one)
As an alternative to the MJI, most of the components of this instrument were designed to correspond to Kohlberg’s stages of moral development. In the example above, a corresponds to Stage 1, b corresponds to Stage 2, c corresponds with Stage 5, d corresponds with Stage 3, e corresponds partially but not with any particular stage, and f corresponds with Stage 4.

Scoring of the SROM is based on respondents’ mean stage level of responses to the eight sociomoral norms. Unlike the SRM that offers two indices, the SROM offers only the SROM maturity scale (SROMS) (without the modal stage score) that is similar to the SRMS but ranges from 100 to 500.

Test-retest studies on the SROM conducted with a two-week interval yielded a correlation of .82 (.76 with age partialed out); the lowest correlation of .70 was from the seventh graders but this is considered acceptable. As part of validating psychometric properties of the SROM, the Cronbach’s alpha was computed as .84 based on the limited sample; this indicated adequate internal consistency of the measure. In addition, the SROM was found to correlate adequately with the SRM (r (81) = .73, p < .001) in a study with 82 subjects aged 11 to 22 (mean age = 14.5 years old). The SROM also correlated adequately with the MJI (r (21) = .66, p < .001), though this result should be interpreted with caution as only 23 subjects aged 13 to 41 (mean age = 20.1 years old) were involved.

To gather further evidence on the SROM’s construct validity, Gibbs et al. (1984) studied correlations between the SROM and variables including age, grade, IQ, SES and social desirability. They found that correlations with age and grade were significant and that with IQ was significant with a larger sample. Though the
correlation with grade was in the .60s, an upward progression was observed in the mean SROMS and the SRMS; this indicated that Kohlberg’s theory could be demonstrated by grade levels. While the evidence gathered supported the reliability and validity of the SROM, Gibbs et al. (1984) expressed that the SROM could not distinguish delinquents from non-delinquents when IQ was not partialled out. They conceded based on the evidence they had collected that the SROM, while a feasible objective recognition measure of moral reasoning, might not be applicable to all adolescent and adult age levels of the population e.g., sixth graders as reading literacy is a prerequisite.

In an attempt to further shorten the SROM, Basinger and Gibbs (1987) developed the Sociomoral Reflection Objective Measure-Short Form (SROM-SF). The SROM-SF is group administrable and more objective than instruments that involve inferences by interviewers and is hence easier to score; it is the “shortest and least demanding. Respondents have to complete a questionnaire comprising two moral dilemmas and 48 moral reasoning justifications in the SROM-SF. The SROM-SF excludes the items that were more verbally complex in the SROM; items related with Stage 5 and partially high stage (pseudo) options were also excluded. Basinger and Gibbs (1987) found that the SROM-SF was reliable and valid specifically with eleventh graders and required about 20 minutes less than the SROM for administration. Nonetheless, as with the SROM, there was a lack of evidence to conclude that the SROM-SF would be applicable to sixth graders and juvenile delinquent adolescents. Given its specific applicability, the SROM-SF did not seem to achieve wide usage and there are limited discussions in the literature.
Following the development of the SROM-SF, Gibbs et al. (1992) developed the *Sociomoral Reflection Measure-Short Form* (SRM-SF) in an effort to shorten and simplify the initial more complex SRM for efficiency. Similar to the SRM, the SRM-SF is a production measure anchored on Kohlberg’s theory. Though shorter and initially touted by Gibbs and colleagues as more group administrable than the SRM, the extent of group administration of the SRM-SF is questionable beyond the classroom context given that interviews still have to be conducted, transcribed, analysed and scored (Brugman et al., 2007).

Respondents for the SRM-SF are required to complete an 11-item questionnaire by circling, for each item, one of the options “very important/important/not important” and explaining in writing why they chose that option. Instead of longer moral dilemmas, short scenarios were used. In doing the questionnaire, respondents will evaluate and justify the importance of seven sociomoral norms: (i) affiliation; (ii) life; (iii) law; (iv) legal justice; (v) truth; (vi) contract; and (vii) property. This is a slight departure from the eight sociomoral norms assessed by the SRM and SROM; “family affiliation” seems to have been combined with “affiliation” and “conscience” has been changed to “truth”. The 11 items are:

1. Think about when you’ve made a promise to a friend of yours. How important is it for people to keep promises, if they can, to friends?
2. What about keeping a promise to anyone? How important is it for people to keep promises, if they can, even to someone they hardly know?
3. How about keeping a promise to a child? How important is it for parents to keep promises, if they can, to their children?
4. In general, how important is it for people to tell the truth?
(5) Think about when you’ve helped your mother or father. How important is it for children to help their parents?

(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?

(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?

(8) How important is it for a person to live even if that person doesn’t want to?

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

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

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

Table 12 presents the sociomoral norms underlying each item.

Table 12. Mapping of SRM-SF items to sociomoral norms

<table>
<thead>
<tr>
<th>Item Number</th>
<th>Sociomoral norm undergirding item</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Contract with friends</td>
</tr>
<tr>
<td>2</td>
<td>Contract with anyone</td>
</tr>
<tr>
<td>3</td>
<td>Contract with children</td>
</tr>
<tr>
<td>4</td>
<td>Truth</td>
</tr>
<tr>
<td>5</td>
<td>Affiliation with parents</td>
</tr>
<tr>
<td>6</td>
<td>Affiliation with friends</td>
</tr>
<tr>
<td>7</td>
<td>Life of a stranger</td>
</tr>
<tr>
<td>8</td>
<td>Life of one self</td>
</tr>
<tr>
<td>9</td>
<td>Property</td>
</tr>
<tr>
<td>10</td>
<td>Law</td>
</tr>
<tr>
<td>11</td>
<td>Legal justice</td>
</tr>
</tbody>
</table>
Scoring is done on respondents’ justifications to probing questions. Similar to the SRM, the SRM-SF provides an index known as the Sociomoral Reflection Maturity Score (SRMS-SF) that ranges from 100 (pure stage 1) to 400 (pure stage 4). Stages 5 and 6 of Kohlberg’s model are excluded as justifications to these stages require higher verbal abilities (Bock, 2008). Gibbs and colleagues opined that the SRMS-SF is comparable to the MMS of the MJI for the first four stages. As an alternative to the SRMS-SF, results can also be presented in 10 bands known as the 10-level Global Stage Index (GSI): (1) 100-125 = GSI stage 1 (2) 126-149 = GSI major-minor transition 1 (2) 150-174 = GSI major-minor transition 2 (1) (4) 175-225 = GSI stage 2 (5) 226-249 = GSI major-minor transition 2 (3) (6) 250-274 = GSI major-minor transition 3 (2) (7) 275-325 = GSI stage 3 (8) 326-349 = GSI major-minor transition 3 (4) (9) 350-374 = GSI major-minor transition 4 (3), and (10) 375-400 = GSI stage 4.

Correlation analyses have been performed to evaluate the reliability and validity of the SRM-SF (see Brugman et al., 2007). As with other derivatives of the SRM, these included inter-rater reliability, test-retest reliability (test-rest took place over an interval of two weeks), split-half reliability, internal consistency using Cronbach’s alpha, and comparisons with the MJI and other variables such as age, SES, IQ, verbal intelligence, and juvenile delinquency. Samples that have provided data for both the SRM-SF and the MJI have included fourth graders (n=40, mean age = 10.02 years old), sixth graders (n=38, mean age = 12.10) and eighth graders (n=37, mean age = 14.03), high school students (n=78, mean age = 17.28) university students (n=42, mean age = 19.19) and delinquents (n=55, mean age = 16.08).
The SRM-SF researchers (see Gibbs et al., 1992) have also reported high inter-rater reliability for this instrument, with correlations between expert raters found to be as high as .99, with a minimum agreement level for trainee raters of .96. Test-retest reliability coefficients for different samples have also been reported to range between .61 and .78, with average coefficients of .88 and .72 with age partialed out. Cronbach’s alphas have also been computed to be around .92 for this instrument. Given the high correlation coefficient values, it can be concluded that the SRM-SF meets minimum criteria in terms of reliability.

Of the four instruments on sociomoral measures, the SRM-SF has been more widely used and can also be applied to a wider age group; it has also stronger evidence for construct validity and reliability compared with the other three (Bock, 2008). Besides, Palmer (2011) stated that the SROM and SROM-SF have proven to have relatively more limited reliability and validity. Bock (2008) highlights that the SRM-SF is unique in that it uses moral behaviours (vignettes) derived from Kohlbergian moral dilemmas instead of the usual lengthy moral dilemmas. Nonetheless, participants are still required to reason in writing their choice of a particular moral behaviour; this could explain why IQ correlated positively with the SRMS-SF as a respondent with a higher IQ score would generally be more articulate in reading and writing. While Bock (2008) states that it has very good to excellent psychometric properties, it does not include Kohlberg’s stages five and six. Both the SRM and SRM-SF have been tested beyond the American context and proven to be applicable though the SRM-SF is easier to score. For example Nilsson, Crafoord, Hedengren and Ekehammar
(1991) applied the SRM to 542 Swedish elementary and high school students (aged 8 to 17 years old) and concluded its applicability to the Swedish context, at the very least for research purposes, considering the acceptable inter-rater reliability that ranged between .83 and .92 and Cronbach’s alpha of .76 that indicated adequate internal consistency. They also found that the SRM was moderately but significantly related to both age and grade, similar to the findings of the SRM researchers. In a separate study, Ferguson, McLernon and Cairns (1994) applied the SRM-SF to a Northern Irish sample and concluded that it is suitable in an Irish or British context. Despite these results, it should be noted here that the scenarios presented in the SRM-SF and the moral dilemmas of the SRM may not be familiar within the Asian context, given that the sample used for validating the SRM and the SRM-SF were not specifically from an Asian context. Thus, cultural differences could still impact the scores.

In a further bid to reduce administration and coding time, Brugman et al. (2007) developed the Sociomoral Reflection Measure – Short Form Objective (SRM-SFO). The SRM-SFO combines both the SRM-SF and SROM-SF, noting the shortcomings of both e.g., the SROM-SF poses problems for younger respondents especially those who have reading difficulties given the moral dilemmas are used in the instrument, and for the SRM-SF, being a production measure, it requires fundamental writing skills and likely pose challenges for administration beyond the size of a classroom. The SRM-SFO comprises ten sets of questions that participants rate and rank in about 15 minutes, much lesser time required than its predecessors. Figure 4 and Figure 5 present example items of the SRM-SFO.
1. Think about when you’ve made a promise to a friend.

I. How important is it for people to keep promises, if they can, to friends?
   □ Very important
   □ Important
   □ Not important

II. If you had to give a reason WHY it is IMPORTANT to keep a promise to a friend if you can, what reason would you give?
   Is this close to a reason you would give?
   □ Yes
   □ No
   □ Not sure

   A. Because your friend may have done things for you, and you need friends
   □ Yes
   □ No
   □ Not sure

   B. Because friendships as well as society must be based on trust.
   □ Yes
   □ No
   □ Not sure

   C. Because otherwise that person won’t be your friend again.
   □ Yes
   □ No
   □ Not sure

   D. Because otherwise you would lose trust in each other.
   □ Yes
   □ No
   □ Not sure

III. Of the reasons given, which one is the closest to the reason you would give?
   □ A  □ B  □ C  □ D

Figure 4. Example item (1) of the SRM-SFO*

11. People are not allowed to force others into having sex.

I. Is it right or wrong that people are \textit{not} allowed to force others into having sex?

- Right
- Usually right
- Sometimes right
- Wrong

II. If you had to give a reason why it is (at least sometimes) right that people are \textit{not} allowed to force others into having sex, what reason would you give?

\textbf{Does this resemble a reason you would give?}

A. Because in a society you have to respect other people's rights, including whether or not to have intimate relations with someone

- Yes
- No
- Not sure

B. Because forcing people to have sex may cause more problems than pleasure

- Yes
- No
- Not sure

C. Because otherwise the other person will turn you in and you will go to jail

- Yes
- No
- Not sure

D. Because it is hard to imagine a more selfish or indecent person than a rapist

- Yes
- No
- Not sure

III. Of the reasons given, which one resembles the one you would give the most?

- A
- B
- C
- D (Choose one)

\textbf{Figure 5. Example item (11) of SRM-SFO*}

The SRM-SFO provides respondents with two scores: (i) the Moral Value Evaluation (MVE) score that measures the level of importance a respondent considers on seven sociomoral norms similar to the SRM-SF i.e., contract, truth, affiliation, life, property, law and legal justice, and (ii) the SRMS that indicates the moral reasoning stage; it ranges from 100 (pure stage 1) to 400 (pure stage 4), similar to that of the SRM.

In validating the SRM-SFO, Brugman et al. (2007) focused on comparing the scores of delinquents and non-delinquents. The sample of non-delinquents involved 305 adolescents (130 male) aged between 11.9 and 16.8 years \( (M = 14.1, SD = 1.1 \text{ years}) \) from six secondary schools of lower, middle and higher tiers. The sample of delinquents that provided data throughout the study involved 151 male adolescents aged between 13 and 22 \( (M = 16.9, SD = 1.4 \text{ years}) \). The study found that the moral maturity of non-delinquents \( (\text{SRMS}_{\text{mean}} = 281, SD = 40.4, \text{range: 170-367}) \) was unexpectedly slightly lower than that of delinquents; this could be explained by the older age of the delinquent group \( (\text{SRMS}_{\text{mean}} = 294, SD = 26.8, \text{range: 225-369}) \). To confirm this unexpected result, Brugman and colleagues reorganised the initial sample for a more comparable non-delinquent and delinquent sample; they selected males below 16 years old and omitted those who have had police contact for the non-delinquent sample and selected 45 delinquents. This updated sample comprised 107 non-delinquents \( (M = 14.0, SD = 1.0) \) and 45 delinquents \( (M = 15.1, SD = .71) \). Nonetheless, even with this comparable sample, the results yielded a higher SRMS and MVE for the delinquent sample, contrary to the finding of the SRM-SF (that delinquents scored lower than non-delinquents).
Further to the comparison between delinquents and non-delinquents, Brugman and colleagues performed a confirmatory factory analysis to confirm the a priori factorial structure of the SRM-SFO and ascertain that it is measuring one general underlying latent construct (see Figure 6 and Figure 7). This was followed by a multi-group confirmatory factor analysis to compare the delinquent and non-delinquent groups. The multi-group analysis had acceptable fit $\chi^2 (70, N = 152) = 75.96, p < .29; CFI = .97; RMSEA = .014$). It is noteworthy that some of the item loadings across both models are small and the loading of item 1 for the delinquent sample is -.04. This could be attributed to the small sample. In addition, the sample size likely yielded a non-significant chi-square statistic. More studies need to be undertaken before confirming the structure of the SRM-SFO.

Given its more recent development and the lack of validity and reliability studies, the SRM-SFO has not been used as widely as the SRM to date. Further, the example item (11) in Figure 5 may be inappropriate or incomprehensible for subjects at a younger age especially if the cultural context where a subject is in does not discuss sexual matters at an early age. Nonetheless, Brugman et al. (2007) concluded that the SRM-SFO can be a promising measure of moral reasoning for adolescents given its format, ease of administration and acceptable level of reliability.
Figure 6. Measurement model for non-delinquent sample
Figure 7. Measurement model for delinquent sample
2.5. THE NEED FOR A NEW MORAL REASONING INSTRUMENT

A summary of all measures reviewed in previous sections is presented in Table 13. The Singapore CCE syllabus recommends a variety of assessment practices for teachers involved in CCE, which include self-assessment, peer assessment, and teacher assessment. Recommended tools across these forms of assessment include checklists, journals, rubrics, behavioural indicators, and holistic report cards. To date, however, there is no known standardised assessment for any aspect of the MOE CCE syllabus. As a result, there are no standardised instruments available to schools for the assessment of moral development in students. The lack of a standardised tool to assess moral reasoning introduces a potential problem in the lack of consistency with which schools may apply and evaluate students’ attainment of the learning outcomes stipulated in the CCE.
### Table 13. Chronological summary of moral reasoning measures

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Type of measure</th>
<th>Number of parallel test forms</th>
<th>What participants have to do</th>
<th>Psychometric properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kohlberg (1958) Moral Judgment Interview (MJI)</td>
<td>Production</td>
<td>Two (Form A and B)</td>
<td>Construct response verbally or in writing to an interview based on a minimum of 21 probing questions per dilemma. There are three dilemmas per test form.</td>
<td>Acceptable levels of reliability and validity</td>
</tr>
<tr>
<td>Rest (1979) Defining Issues Test (DIT)</td>
<td>Recognition</td>
<td>One</td>
<td>Select response (i.e., rate and rank) to 12 issue statements related to the initial decision to a moral dilemma. There are six dilemmas (three are Kohlbergian dilemmas).</td>
<td>Acceptable levels of reliability and validity</td>
</tr>
<tr>
<td>Page &amp; Bode (1980) Ethical Reasoning Inventory (ERI)</td>
<td>Recognition</td>
<td>One</td>
<td>Select response (i.e., multiple-choice) to six moral dilemmas. There are six options each corresponding to a “yes” or “no” option for each dilemma.</td>
<td>Acceptable levels of reliability and validity</td>
</tr>
<tr>
<td>Gibbs et al. (1982) Sociomoral Reflection Measure (SRM)</td>
<td>Production</td>
<td>Two (Form A and B)</td>
<td>Construct responses to two moral dilemmas based on eight probing questions.</td>
<td>Acceptable levels of reliability and validity</td>
</tr>
<tr>
<td>Gibbs et al. (1984) Sociomoral Reflection Objective Measure (SROM)</td>
<td>Recognition</td>
<td>One (modelled after Form A of the SRM)</td>
<td>Select response (i.e., rate and rank) 16 multiple-choice arrays.</td>
<td>No acceptable validity and reliability for sixth graders and juvenile delinquents (Basinger &amp; Gibbs, 1987)</td>
</tr>
<tr>
<td>Basinger and Gibbs (1987) Sociomoral Reflection Objective Measure – Short Form (SROM-SF)</td>
<td>Recognition</td>
<td>One</td>
<td>Select response (i.e., rate and rank) to two moral dilemmas.</td>
<td>No acceptable validity and reliability for sixth graders and juvenile delinquents (Basinger &amp; Gibbs, 1987)</td>
</tr>
<tr>
<td>Gibbs et al. (1992) Sociomoral Reflection Measure – Short Form (SRM-SF)</td>
<td>Production</td>
<td>One</td>
<td>Construct responses to five short moral vignettes based on 11 questions.</td>
<td>Acceptable levels of reliability and validity</td>
</tr>
<tr>
<td>Rest et al. (1999b) DIT2</td>
<td>Recognition</td>
<td>One</td>
<td>Select response (i.e., rate and rank) to 12 issue statements related to the initial decision to a moral dilemma. There are five moral dilemmas.</td>
<td>Acceptable levels of reliability and validity</td>
</tr>
<tr>
<td>Brugman et al. (2007) Sociomoral Reflection Measure – Short Form Objective</td>
<td>Recognition</td>
<td>One</td>
<td>Select response to 10 dilemma free items</td>
<td>Inconclusive</td>
</tr>
</tbody>
</table>
Despite the substantial body of research evidence which supports the validity of the instruments discussed thus far, for several reasons, none would be suitable for assessing development on a broad-scale basis within the Singapore CCE curriculum. First, while the MJI does provide an extensive evaluation of students’ moral reasoning, this test is individually administered, and extremely time-consuming and complex to score (Miller, 2007). Teachers in schools who are charged with assessing, in some cases, hundreds of students concurrently, would not find this a feasible instrument to use. Teachers would also have to go through extensive training on the MJI scoring protocols and systems to minimise disparities. The same issues would also apply to other production measures such as the SRM and SRM-SF, and for these reasons, the latter instruments are also not suitable for use within this context.

Second, while various established instruments have been used with children across ages and cultures, students in the Singapore secondary context may not relate to the scenarios presented within these instruments. Most focus on issues that will be unfamiliar to students in their day-to-day lives. As an example, while time efficient, the SRM-SFO has an item that to younger secondary school students may have difficulty relating with, as depicted in Figure 5.
III. People are not allowed to force others into having sex.

I. Is it right or wrong that people are not allowed to force others into having sex?
   - Right  □ Usually right  □ Sometimes right  □ Wrong

II. If you had to give a reason why it is (at least sometimes) right that people are not allowed to force others into having sex, what reason would you give?

   Does this resemble a reason you would give?
   A. Because in a society you have to respect other people’s rights, including whether or not to have intimate relations with someone
      □ Yes  □ No  □ Not sure
   B. Because forcing people to have sex may cause more problems than pleasure
      □ Yes  □ No  □ Not sure
   C. Because otherwise the other person will turn you in and you will go to jail
      □ Yes  □ No  □ Not sure
   D. Because it is hard to imagine a more selfish or indecent person than a rapist
      □ Yes  □ No  □ Not sure

III. Of the reasons given, which one resembles the one you would give the most?
    □ A  □ B  □ C  □ D  (Choose one)

Figure 5
Third, some hypothetical moral dilemmas used in these instruments may be inappropriate for students in the Singapore secondary context given their complexity and that they are sometimes lengthy and difficult to comprehend. Gibbs, Basinger and Fuller (1992) conceded that: (i) the complexity of the SRM-SF, a supposedly short moral reasoning measure comprising two moral dilemmas and 48 moral reasoning justifications “including fairly sophisticated ones” (p.35); (ii) the SRM has a format that sometimes makes it confusing for younger students. Though a shorter alternative of the SRM, the SRM-SF that is without lengthy moral dilemmas, may make it more accessible to students at this age, this form requires students to justify their selected responses in writing. This
aspect could then introduce construct-irrelevant sources of variance, via differences in writing ability.

An extensive literature search did yield one instrument that has been used to assess moral reasoning in the Singapore context. Soh (1987) developed the *Test of Moral Values* (TMV) for Singapore students in response to a call for a more organised moral education in Singapore. Unlike the previously discussed moral reasoning instruments that are anchored on established theoretical models (e.g., Kohlberg’s theory of moral development), the TMV comprises 24 items each anchored on a different moral value recommended in a 1979 Singapore government report on moral education by the then Communications Minister and Acting Minister for Culture, Ong Teng Cheong. The 24 moral values recommended by the report presented in Table 14. Figure 8 presents an example item of the TMV.

Each option in each item within the TMV corresponds to a given category. Option *(a)* corresponds to the “self” category; Option *(b)* corresponds to the “social/peer influence” category; and Option *(c)* corresponds to the “moral value” category. Despite the fact that there are three categories within the instrument, these categories cannot be mapped to Kohlberg’s model of pre-conventional, conventional and post-conventional moral stages. Hence, the moral values upon which the items are anchored are not grounded in any of the established theoretical frameworks discussed in Chapter One. Furthermore, the validation work on the TMV has been very minimal (see Soh, 1987). Hence, the TMV, at this point, remains non-validated for use in this context.
### Table 14. 24 moral values in the TMV*

<table>
<thead>
<tr>
<th>Number</th>
<th>Moral value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Filial piety</td>
</tr>
<tr>
<td>2</td>
<td>Spirit of inquiry</td>
</tr>
<tr>
<td>3</td>
<td>Self-discipline</td>
</tr>
<tr>
<td>4</td>
<td>Humility</td>
</tr>
<tr>
<td>5</td>
<td>Perseverance</td>
</tr>
<tr>
<td>6</td>
<td>Trust</td>
</tr>
<tr>
<td>7</td>
<td>Loyalty (to School)</td>
</tr>
<tr>
<td>8</td>
<td>Forgiveness</td>
</tr>
<tr>
<td>9</td>
<td>Social Work</td>
</tr>
<tr>
<td>10</td>
<td>Caring for aged</td>
</tr>
<tr>
<td>11</td>
<td>Courtesy</td>
</tr>
<tr>
<td>12</td>
<td>Respect for law</td>
</tr>
<tr>
<td>13</td>
<td>Neighbourliness</td>
</tr>
<tr>
<td>14</td>
<td>Respect for religion</td>
</tr>
<tr>
<td>15</td>
<td>Respect for cultures</td>
</tr>
<tr>
<td>16</td>
<td>Generosity</td>
</tr>
<tr>
<td>17</td>
<td>Pride in country</td>
</tr>
<tr>
<td>18</td>
<td>Loyalty (to Nation)</td>
</tr>
<tr>
<td>19</td>
<td>Commitment</td>
</tr>
<tr>
<td>20</td>
<td>National defence</td>
</tr>
<tr>
<td>21</td>
<td>Justice</td>
</tr>
<tr>
<td>22</td>
<td>Equality</td>
</tr>
<tr>
<td>23</td>
<td>National survival</td>
</tr>
<tr>
<td>24</td>
<td>National prosperity</td>
</tr>
</tbody>
</table>

* As recommended by the 1979 report on moral education

---

**When Kuo Ming has to go home late, he makes sure he tells his parents. Why?**

- a. He does not want to be scolded by them.
- b. His friends have told him to do so.
- c. He does not want his parents to worry about him.

---

**Figure 8. Example item in the TMV.**

---

85
Based on the review conducted here, it was concluded that none of the existing instruments would provide a practical means for Singapore teachers to assess moral reasoning within the CCE curriculum. While some instruments, such as the MJI and the DIT/DIT2 are supported by substantial research evidence, these instruments may not draw upon scenarios that would be familiar to Singapore secondary students. The demands of administration associated with the MJI would also be prohibitive within this context. Other instruments reviewed were either not supported by strong validation evidence (e.g., the TMV), or suffered from similar issues to those identified with the MJI and DIT/DIT2. In light of this, the goal of the study reported in subsequent chapters was to develop a measure that was suitable for assessing Singapore secondary students’ moral reasoning development with the CCE curriculum.
Chapter 3. Method

This chapter describes the procedures used to develop and validate the Moral Reasoning Questionnaire (MRQ), which was designed specifically to assess Singapore secondary students’ moral reasoning development based on Kohlberg’s stage-based model. The MRQ was validated in a series of three major stages, which corresponded with addressing the first three criteria stipulated by Messick (1993, 1995) in his elaboration of his unified concept of validity.

3.1. Messick’s Unitary Concept of Validity

Messick (1993, 1995) defined validity as an integrated evaluation of the extent to which “empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessment” (p. 1). In studying the foundations of validity for psychological assessments, Messick proposed that the traditional, three-part framework on validity (i.e., content, criterion, and construct validities) espoused in the original 1955 Standards for Educational and Psychological Testing, as well as all associated concepts that have emerged in the field since that time (e.g., face validity, convergent validity, concurrent validity, and discriminant validity) encouraged a fragmented view of the validation process. Further, Messick proposed that the three-part framework was incomplete, because it did not account for “evidence of the value implications of score meaning as a basis for action and of the social consequences of test use” (p.1).
Elaborating upon his unified concept of validity, Messick (1993, 1995) also stipulated five forms of evidence that should be gathered by test developers in evaluating the validity of their measures: (i) evidence related to the test content; (ii) evidence related to response processes required by the test; (iii) evidence related to the internal structure of the test; (iv) evidence related to the associations between the test and other variables; and (v) evidence related to the consequences of testing. Messick noted, however, that these were not to be interpreted as separate ‘types’ of validity, just as different forms of evidence that can contribute to the overall evaluation of the validity of a measure.

Messick’s views on validity have now been adopted widely within the educational and psychological measurement communities. Of particular note, Messick’s framework formed the basis for many of the revisions that now appear within the most recent Standards for Educational and Psychological Testing manual published by the American Educational Research Association (AERA), American Psychological Association (APA) and the National Council on Measurement in Education (NCME). Through their adoption in this manual and in other education measurement fora, Messick’s views on validity have radically altered the ways in which validity is conceptualised and evaluated in the context of educational and psychological testing. In particular, the 2014 Standards for Educational and Psychological Testing manual encourages test developers to focus on various sources of relevant evidence to support the score interpretations yielded by given measures.
Applying these notions to the current case, for the MRQ to be deemed an acceptable measure of moral reasoning, the scores generated by the MRQ should prompt meaningful interpretations about a respondent’s level of development in moral reasoning, rather than any other attributes of that respondent. In the validation of the MRQ conducted in this study, evidence on the content of the MRQ, on its responses processes, and on its internal structure were obtained. Owing to the constraints under which the study was conducted, it was not possible to gather evidence on associations between the MRQ and other variables. This was due to the fact that the participating schools were reticent to allow the administration of more than one instrument in connection with this research. Owing to the timeframe of this research, it was also not possible to gather evidence related to the consequences of testing, because this criterion can only be evaluated after the test has been adopted for use. Thus, the validation conducted here addressed three of the five criteria stipulated by Messick (1993, 1995).

3.2. PARTICIPANTS

In all, there were three groups of participants within the study; all the groups were based on availability sampling. The first group comprised experienced teachers from the Singapore secondary school system, who participated in the initial content validation of the MRQ. Participants in this group included: (i) two officers from the Singapore MOE (one was a male Assistant Director from the Student Development Curriculum Division, with more than 15 years of experience in this role - this division plans and oversees all CCE related policies
and curriculum in schools; the other was a female Guidance Officer with more than five years of experience in handling students ‘at risk’, in particular, those with behavioural issues; (ii) a female Head of the CCE Department of a Singapore secondary school, with more than 30 years of experience managing students; (iii) A female Deputy Head of the CCE Department from a Singapore secondary school, with more than five years of experience in her role; (iv) a female Singapore secondary school Science Senior Teacher cum Year Head, with more than 20 years of experience in schools - she oversaw students’ ‘holistic’ development by working with form teachers to attend to students' socio-emotional and pastoral needs and hence cultivate values; (v) a female Head of Physical Education and Co-Curricular Activities in a Singapore secondary school, with more than 10 years of experience in her role; and (vi) a Singapore secondary school English Senior Teacher cum Assistant Year Head (female), with more than 10 years of pastoral care and teaching experience. This group was selected to ensure that a diverse range of views was obtained on the content of the MRQ, before it was trialled with students.

Two groups of Singapore secondary students also participated in the study. Group A was involved in Stage 2 of the research, and comprised five students selected to participate in a preliminary evaluation of the response processes associated with completing the MRQ. At the time of the study, one of these participants had completed secondary 2, and was 14 years old; one had completed secondary 3 and was 15 years old; and the remaining three had completed secondary 4 and were 16 years of age. All five participants were from the express educational stream in Singapore, and attended different schools.
These included government, government-aided, independent, co-ed and all-boys schools. All except one of the secondary 4 students were boys.

The second group of participating students, Group B, was involved in Stage 3 of the research. This group comprised a large sample of secondary level students from Singapore schools. Access to the schools was granted by the Ministry of Education in September 2015 (see Appendix A1). Six schools (see Table 15) with different profiles were invited via email to participate in the development of the instrument (see Appendix A2). Of these schools, however, only three ultimately agreed to participate in the research, as shown in Table 15.

Table 15. Invited schools and their responses

<table>
<thead>
<tr>
<th>School</th>
<th>Profile / Status</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>School CM</td>
<td>Government-aided / co-ed / autonomous</td>
<td>Principal did not respond to invite.</td>
</tr>
<tr>
<td>School CY</td>
<td>Government-aided / co-ed</td>
<td>Principal did not respond to invite.</td>
</tr>
<tr>
<td>School M</td>
<td>Government / co-ed</td>
<td>Principal agreed to support the study.</td>
</tr>
<tr>
<td>School P</td>
<td>Government-aided / girls' mission school / autonomous</td>
<td>Principal agreed to support the study.</td>
</tr>
<tr>
<td>School T</td>
<td>Government / co-ed / autonomous</td>
<td>Principal said school could not support the study.</td>
</tr>
<tr>
<td>School X</td>
<td>Government-aided / co-ed / autonomous</td>
<td>Principal agreed to support the study.</td>
</tr>
</tbody>
</table>

Across the three schools that eventually agreed to participate in the research (schools M, P and X), the Group B sample comprised 670 Singapore secondary school students (497 female, 173 male). The participants ranged in age from 12 to 18 years ($M = 14.24$, $SD = 1.30$ years) at the time of the study. Within the sample,
36.7% \((n = 246)\) of the students were drawn from school M; 44.2% \((n = 296)\) from school P; and the remaining 19.1% \((n = 128)\) were drawn from school X.

The three participating schools were diverse, representing different educational streams (i.e., express, normal – academic and normal - technical) and educational levels (i.e., secondary one to four). Of the 670 participants, 79.4% \((n = 532)\) were from the express stream, 16.3% \((n = 109)\) from the normal – academic stream, and 4.3% \((n = 29)\) from the normal – technical stream. In terms of education levels, 17.6% \((n = 118)\) of the participants were in the secondary one level at the time of the study; 27.8% \((n = 186)\) in secondary two; 28.4% \((n = 190)\) in secondary three; and 26.3% \((n = 176)\) were in secondary four.

3.3. Instrument Design

3.3.1. Item stems

Based on a thorough critical review of the literature on existing instruments to assess moral reasoning levels, a decision was made to retain the use of moral dilemmas as item stems (i.e., the ‘prompt’ questions) within the MRQ. A large pool of items was generated initially, with the goal of providing meaningful differentiation across Kohlberg’s six stages of moral development (i.e., pre-conventional, conventional and post-conventional). The Kohlberg model, as noted previously, was adopted here in light of its strong alignment with the stipulated outcomes of the CCE curriculum.
A moral dilemma is a situation, sometimes hypothetical, in which moral reasoning is relevant for making a decision. In moral dilemmas, an individual is asked to make a judgement on the action that he or she, or a third party, should take based on his or her moral reasoning processes (Sinnott-Armstrong, 1999; McConnell, 2014). There are several taxonomies that can be used to classify moral dilemmas based on their attributes. One classification, suggested by Vallentyne (1989) posits two distinct types of moral dilemma: (i) *obligatory* moral dilemmas, in which more than one action is obligatory; and (ii) *prohibitive* moral dilemmas, in which no feasible action is permissible in a situation. Of the two, obligatory dilemmas are more likely to be familiar to secondary school students, because this format is more typically used in education-based contexts.

Another taxonomy that has some relevance for secondary school students is based on the classification by Wark and Krebs (2000), who suggested that there are three main types of moral dilemma: (i) those that are antisocial in nature; (ii) those that are prosocial in nature; and (iii) those that relate to resisting or complying with social pressure in dilemma situations. Dilemmas antisocial in nature generally prompt the respondent to consider actions that should be taken when individuals break rules or behave dishonestly. Dilemmas prosocial in nature typically involve asking the respondent to consider situations that call upon prosocial behaviours such as helping a friend or deciding how best to respond when one wants to be loyal to two friends/relatives. Dilemmas related to social pressure generally involve asking respondents to decide how they, or other individuals, should act when others pressure them to act in a way that contradicts sound moral values (e.g., stealing). The latter types of dilemmas can
also involve prompting an individual to think about how he or she, or others, should respond in the face of parental or authority-grounded pressure.

Christensen, Flexas, Calabrese, Gut and Gomila (2014) noted that moral dilemmas have become a standard methodology for research on moral judgment. Evidently, most instruments to assess moral judgment or reasoning are based either partly or wholly on items that rely on moral dilemmas as prompts. One may argue that moral dilemmas are often unrealistic, and hence do not indicate moral reasoning accurately. This can, however, be overcome by using more realistic and familiar dilemmas. In fact, in one early study, Sumfrer and Butter (1978) found no differences between college students’ levels of moral reasoning to hypothetical and actual moral dilemmas.

Based on the review conducted here of instruments that use moral dilemmas as a stimulus to assess moral reasoning, obligatory moral dilemmas were selected for use in the MRQ, incorporating scenarios that were prosocial, anti-social and social pressure in nature. Scenarios were developed with the goal of ensuring that these would be familiar and realistic to the intended respondents (i.e., Singapore secondary school students), to overcome any problems associated with the ‘artificial’ nature of the dilemmas.

3.3.2. Response options

Fabbris (2013) classified five methods for data collection in questionnaire situations: (i) ranking of items; (ii) picking the best or worst item; (iii) partitioning a fixed total among the items; (iv) rating each item; and (v) paired comparisons
of all distinct pairs of items. He highlighted that rating approaches, such as Likert scale ratings, often attracted less fatigue from respondents in comparison to ranking and paired comparisons, with paired comparisons requiring the most from respondents. Despite this, the Likert scaling approach was not used in this case, because the nature of the response options typically used in Likert-based scales can introduce interpretational challenges.

Ranking requires respondents to order items based on some kind of identified attribute. Jacoby (2011) posed that ranking a set of values according to their subjective importance has “generally been regarded as the “gold standard” in obtaining empirical representations of individual value structures” (p.2). While ranking provides a convenient representation of respondents’ choices, Clawson and Vinson (1978) highlighted that by using ranking, “equally attractive values are forced into separate rankings” and “wide gaps in preference are treated as no different from minuscule gaps” (p.398). Further, Alwin and Krosnick (1985) stated that “rankings are often difficult and taxing for respondents, demanding considerable cognitive sophistication and concentration” (p.536). Noting its benefits and disadvantages, ranking was selected for use in developing the MRQ, but care was taken to minimise sophistication and cognitive demands. Items were worded with simple language and in short sentences in an attempt to minimise construct-irrelevant sources of variance in the responses.

Based on a comprehensive review of the literature on response formats used in assessing moral reasoning, a decision was made to use a two-tier response format (see Appendix B). Respondents first read the moral dilemma posed in each item.
In Tier One of the response format, the respondent is required to select either of the “action” options posed (see Section 10a of Figure 9). The respondent is then presented with “moral reasoning” options on the subsequent screen depending on which option was selected in tier-one (see Section 10b1 and 10b2 of Figure 9). In sections 10b1 or 10b2 of Figure 9, the respondent is required to rank the options in order of importance to him/herself. Each of the options in 10b1 and 10b2 was designed to correspond with one of Kohlberg’s main stages of moral development (i.e., pre-conventional, conventional and post-conventional).

Further to Kohlberg’s model, the common four strands proposed in Table 9 were considered in designing the items. Taking Figure 9 as an example, as a student reads the item stimulus (10a), he or she would first have to think about possible actions by sensitising him/herself with others (strand one), and how his/her actions would impact others’ emotions (strand two). There are two possible options – that is, to report or not to report Jane to the teacher. Upon selecting the option in 10a, the respondent will be shown 10b1 or 10b2. It is in 10b1 or 10b2 that the respondent has to prioritise the action to take based on self-motivation, values and knowing all the facts presented in the stimulus (strand three). Following this, the respondent has to rank the options in 10b1 or 10b2. By doing so, he or she is demonstrating an appreciation of the consequences and undergirding rationale for each action, and taking steps to prioritise each action (strand four).
10a. The class beside yours took the same Science test 2 hours ago. Over recess, you heard your friend Jane asking what the questions were. Even though this is just a class test, Jane says that she needs whatever help to boost her failing results. Jane normally sits beside you during a test. You would...

- report Jane to the teacher.
- not report Jane to the teacher.

10b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report Jane to the teacher. You would...

- report Jane to the teacher as you might be given a zero if your teacher found out that you knew but didn’t report the matter.
- report Jane to the teacher as the school does not allow cheating.
- report Jane to the teacher as it is more important to ensure fairness in the test for everyone.

10b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report Jane to the teacher. You would...

- not report but ask Jane what the questions were and discuss the answers with her so that both she and you would benefit and she would value you as a friend more.
- not report but ask Jane to share the questions with the class so that everyone in the class can do well.
- not report Jane as this could be a morale booster for her to work harder in future tests.

Figure 9. Example item with vignette and corresponding options
3.3.3. Item scoring
As the MRQ is designed to assess moral reasoning and not moral action, neither of the selected “action” options (i.e., Tier One responses) are scored. Scoring for Tier Two responses is presented in Table 16. It should be noted that the scores do not imply morality as such - the scores indicate the level of moral reasoning based on Kohlberg’s model of moral development. Thus, a respondent with a higher overall score should not be deemed ‘more moral’ than one with a lower overall score. However, as posited by Rest (1979), it is assumed that a respondent who scores higher on the MRQ is at a higher level of moral reasoning than those who obtain lower scores on the instrument.

Table 16. Scoring matrix of two-tier items

<table>
<thead>
<tr>
<th>Rank order (levels of moral judgment)</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Post-conventional (2) Conventional (3) Pre-conventional</td>
<td>5</td>
</tr>
<tr>
<td>(1) Conventional (2) Post-conventional (3) Pre-conventional</td>
<td>4</td>
</tr>
<tr>
<td>(1) Post-conventional (2) Pre-conventional (3) Conventional</td>
<td>3</td>
</tr>
<tr>
<td>(1) Conventional (2) Pre-conventional (3) Post-conventional</td>
<td>2</td>
</tr>
<tr>
<td>(1) Pre-conventional (2) Post-conventional (3) Conventional</td>
<td>1</td>
</tr>
<tr>
<td>(1) Pre-conventional (2) Conventional (3) Post-conventional</td>
<td>0</td>
</tr>
</tbody>
</table>

3.4. Stage 1 Data Collection Procedures: Content Appropriateness of Items - Expert Judgment

Thirty possible items were prepared in consultation with the thesis supervisor. Care was taken to avoid gender and race or religion bias when writing the items. Polysyllabic words were avoided in the moral dilemma vignettes. To the extent
possible, sentences in the items averaged between 15 and 20 words, as recommended by Cutts (2013) in the Oxford Guide to Plain English.

Over a period of two months from February to March 2015, a panel comprising experts in the field of CCE in Singapore schools and the MOE was invited to critique the items and feedback on the content appropriateness of the items (see Appendix A3). Questions for the panel included:

(i) Is the content of each of the 30 items appropriate for 13 to 17 year old students with respect to moral reasoning?

(ii) Will the majority of 13 to 17 year old students be able to understand the language structure and meaning of the moral dilemma and options in each item?

(iii) Are the options for each item clear and comprehensible with respect to moral reasoning?

(iv) Are the instructions on how to respond to the items clear?

(v) Are the items gender, race or religion biased?

A document shown partly in Figure 10 was printed for the panels’ convenience.

In your review, please circle the options that you think are appropriate and provide comments for improvement. For convenience, a table with the options for selection will be appended at the end of each item. An example is shown below:

<table>
<thead>
<tr>
<th>Item 9</th>
<th>The language used in this item is suitable for Sec 1 to 4/5 students.</th>
<th>The language used in this item is not suitable for Sec 1 to 4/5 students.</th>
<th>Comments/Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Students may experience the scenario given in this item.</td>
<td>Students will never experience the scenario given in this item.</td>
<td>Comments/Recommendations</td>
</tr>
<tr>
<td></td>
<td>This item addresses moral reasoning issues.</td>
<td>This item does not address moral reasoning issues.</td>
<td>Comments/Recommendations</td>
</tr>
<tr>
<td></td>
<td>Students will likely understand this item (Item is favourable).</td>
<td>Students will likely not understand this item (Item is not favourable).</td>
<td>Comments/Recommendations</td>
</tr>
</tbody>
</table>

Other comments/recommendations:
- e.g., Options 4 and 5 will be difficult for students to differentiate. Suggest a change of words for option 3 etc.

Figure 10. Document for judges to score and comment

Following the two-month review period, items were scored and tabulated based on the contrasting columns in Figure 10 (e.g., students may experience the
scenario given in this item, coded as 1, vs. students will never experience the scenario given in this item, coded as 0). While most of the items had a high score, attention was paid also to all comments received, and in particular, to items that were poorly scored (i.e., an average score of 2 or lower). To minimise construct-irrelevant sources of variance that might advantage/disadvantage certain groups of students who were stronger or weaker in English, or those with sensitive family or personal issues, certain items were refined in light of the feedback received. For example, item 24 originally described a brother asking the respondent for money for an urgent surgery; two of the panel experts opined that it would be unlikely that a brother would ask a secondary school student for money for a surgery. Hence, this item was refined as a brother asking the respondent for money to buy textbooks. Following the refinements and in agreement with the expert panel, all 30 items were deemed to have adequate content appropriateness, and ready for stage 2 of the validation process.

3.5. STAGE 2 DATA COLLECTION PROCEDURES: EVIDENCE BASED ON RESPONSE PROCESSES

The processes used in Stage 2 were designed to evaluate the validity of the MRQ in terms of the second of Messick’s (1993) criteria (i.e., evidence related to response processes required by the test). In Stage 2, parental consent was sought for the five students in Group A (see Appendix A4) to complete the 30 refined MRQ items. Each participant was interviewed individually throughout the process of completing the items. This form of validation, based on the notion of
‘cognitive interviews’, is now widely used in the validation of survey questions (Beatty & Willis, 2007). The cognitive interview method is a method of studying the “mental processes one uses when completing a task such as solving a mathematics problem or interpreting a passage of text” (Zucker, Sassman & Case, 2004, p.2). Thus, this process was used here to provide insight into the mental processes used by respondents in completing each MRQ item.

Each interview took approximately one hour. The processes used in these interviews followed the four-step model recommended Ryan, Gannon-Slater, and Culbertson (2012) (see Figure 11): (i) question comprehension, where clarity and the intended interpretation of the questionnaire instructions and items can be sought, (ii) information retrieval, where clarity on how a participant recalls information to respond to the item is sought, (iii) judgment and estimation, where understanding on how participants arrive at a decision is sought, and (iv) documenting a response – whether a participant can match his/her response to the response-options provided by the items.

Verbal probing was used to gather evidence based on the cognitive interview model. Examples of such probes included: (i) What did you understand from the scenario? (ii) Did you have any difficulty reading and understanding the scenario? (iii) Did you have any difficulty ranking the options? (iv) Were the options provided sufficiently distinct for you to distinguish and rank them?

Applying the model in Figure 11 for stage 2 along with verbal probing, participants first ranked the items as per the written instructions in the MRQ.
They then justified verbally why they chose those rankings and whether their responses reflected their thoughts.

![Cognitive Interview Model](image)

**Figure 11.** Cognitive interview model by Ryan et al. (2012)

With the findings from the cognitive interviews, evidence of the extent to which the rankings represent the level of moral reasoning that the participant had used in responding to each item was gathered. The cognitive interviews revealed that the rank order of the options in Tier Two were distinguishable by the respondents, and that the respondents did not have difficulties in ordering the options or choosing between the two “action” options presented in Tier One of each item. The interviewees also did not have any difficulty understanding the scenarios. No difficult words were highlighted. The evidence from the
interviewees suggested that secondary school students would be able to respond to most, if not all, of the items without much difficulty.

**3.6. Stage 3 Data Collection Procedures: Evidence Based on the Internal Structure of the Test**

In Stage 3, 670 students from the three schools that agreed to support this study completed all 30 items in the MRQ. Students from two of the three schools completed the MRQ via the survey software platform Qualtrics™. Students from the remaining school completed a paper-based version of the MRQ due to logistical and administrative constraints identified by the school. Responses from the paper-based MRQ were input to the main dataset from Qualtrics. All of the students completed the 30 items during curriculum time and within 40 minutes. Parental consent was sought before each of these students attempted the MRQ (see Appendix A5). Before the students did the questionnaire, teachers who helped to administer were given an information sheet and were asked to read a standardised set of instructions to the students (see Appendix A6). Students were reminded that they could withdraw at any time and that their responses should be as honest as possible, as results of the MRQ would not affect their school results in any way. After all data were gathered from the schools, these were then entered into a single file for analysis purposes.
Chapter 4. Results

This chapter presents the results of all analyses performed on the *Moral Reasoning Questionnaire* (MRQ) developed. Initially, for the purposes of the analysis, the total sample was split randomly into two halves, stratifying for school and class. Odd-numbered cases were then used for purposes of all exploratory analyses performed, while even-numbered observations were used to confirm the results of the analyses performed on the odd-numbered cases.

The analysis of the structure of the MRQ relied on three main procedures. Exploratory factor analysis (EFA) was first used to provide a preliminary assessment of the structure of the MRQ, using the odd-numbered cases within the dataset. Confirmatory factor analysis (CFA) was then used to determine whether the conclusions reached on the basis of the odd-numbered cases could be replicated with the even-numbered dataset. Both of these statistical methods are used widely in education and the social sciences to establish the psychometric properties of instruments based on classical test theory (CTT). The MRQ’s psychometric properties were then investigated further using Rasch-based (modern test theory, or MTT) analysis methods. Conclusions about the internal structure of the instrument thus relied on triangulating the results obtained using all three of these methods.
4.1. EXPLORATORY FACTOR ANALYSIS

The Statistical Analysis System (SAS) Version 9.4, an analytics and data management software, was used to perform all of the EFA analysis procedures used in the research. As noted previously, EFA was used to provide a preliminary assessment of the factorial structure of the MRQ, and hence, of the dimensionality of the construct measured by the MRQ. Prior to conducting the EFA, all assumptions underlying the use of this procedure were evaluated thoroughly to ensure that its use was tenable in this case.

First, in terms of sample size, there were 335 odd-numbered observations available for use in the EFA, and there were no missing data within the set. Recommendations for minimum sample sizes and/or case-to-variable ratios for EFA (e.g., Gable & Wolf, 1993; Mundfrom, Shaw & Ke, 2005) vary so significantly within the education research literature that many scholars have argued that common ‘rules of thumb’ for EFA sample sizes are “not valid or useful” (MacCallum, Widaman, Zhang & Hong, 1999, p.96). In a synthesis of the literature related to minimum EFA sample sizes/case-to-variable ratios, Beavers et al. (2013) noted that the largest overall minimum sample size recommended in the papers they reviewed was $n=300$, with a minimum case-to-variable ratio of 5:1. Given that both of these requirements were met within the current study, the use of EFA based on this dataset was deemed to be tenable.

Second, tests for univariate and multivariate normality on the MRQ total and item scores within the EFA dataset produced satisfactory results. For example, using the MRQ total scores, a very modest level of negative skew (-0.74) and a
modest level of kurtosis (0.65) was obtained. Given that these co-efficients fall well within accepted thresholds (skewness < 2 and kurtosis < 7: see West, Finch & Curran, 1995), the assumption of normality was deemed to be tenable in this case. Similar co-efficients were generated on the basis of the individual item score distributions within the EFA dataset. All other assumptions associated with the use of EFA, including linearity and the absence of outlying scores, were also evaluated, with results confirming that the use of EFA in this case was tenable.

A maximum likelihood (ML) procedure was used as the factor extraction method within the EFA. According to Fabrigar, Wegener, MacCallum and Strahan (1999), as long as multivariate normality is not severely violated, ML “allows for the computation of a wide range of indexes of the goodness of fit of the model [and] permits statistical significance testing of factor loadings and correlations among factors and the computation of confidence intervals” (p. 277). As noted previously, there were no significant violations of the normality assumption based on the preliminary investigations performed. However, to ensure that the very modest level of non-normality within the data distributions did not affect the factor extraction process, an unweighted least squares (ULS) extraction (an extraction method that does not assume normal data distributions), was also performed initially for the purposes of comparison. As all of these analyses produced similar results to those obtained through the ML procedure, the ML results were retained for interpretation within this chapter.

To determine whether the sample size of 335 was sufficient for the extraction of factors, Kaiser’s (1970) measure of sampling adequacy (MSA) was also
computed. The MSA essentially indicates how much smaller the partial correlations are in comparison to the original correlations, and how substantial a portion of these correlations the extracted factor(s) would account for. According to Hutcheson and Sofroniou (1999), Kaiser (1974) recommends values greater than 0.5. As a rule of thumb, values between 0.5 and 0.7 are considered ‘mediocre’; values between 0.7 and 0.8 are considered ‘good’; values between 0.8 and 0.9 are deemed ‘great’; and values above 0.9 are considered ‘superb’. A MSA value of at least 0.5 will mean that the data can be meaningfully analysed by factor analysis. In this case, the MSA index was found to be .90 (for both ML and ULS methods), which is considered ‘great’, hence confirming that the use of EFA was appropriate in this case. Bartlett’s Test of Sphericity was also significant ($\chi^2 (435) = 2315.37, p < .001$), further confirming that the data were suitable for EFA.

Based on Kaiser’s (1958) eigenvalue rule, which suggests retaining factors with eigenvalues one or more, the initial analysis (without specifying the number of factors to be extracted or rotated) yielded a single strong factor that explained 76.72% of the total score variance using ML estimation, and 75.88% using ULS estimation, for the entire set of variables. Visual inspection of the scree plots using either the ML or ULS estimation methods (see Figure 12 and Figure 13) also provided warrant for the view that a single strong factor accounted for most to the variance in the MRQ scores.
Figure 12. Scree plot and variance explained figures by ML estimation

Figure 13. Scree plot and variance explained figures by ULS estimation

To minimise the risk of under/over-estimation of factors and hence dimensionality in the interpretation of the MRQ results, a parallel analysis was then conducted with a Monte Carlo simulation of 1000 simulated datasets. Essentially, parallel analysis is a simulation method used to determine the
number of factors to retain in EFA. It was developed based on Cattell’s scree plot to minimise issues related to factor indeterminacy (Franklin, Gibson, Robertson, Philip, Pohlmann, & Fralish, 1995). While it is a less known procedure to determine dimensionality, it is recommended as the best procedure to assess the true number of factors as the Kaiser criterion may under/over-estimate the true number of factors (Basto & Pereira, 2012; Ledesma & Valero-Mora, 2007). The graphical output of the parallel analysis for this study is presented in Figure 14. Hence, the results of the EFA and of the parallel analysis were convergent in suggesting that the MRQ represents a unidimensional measure. The factor patterns obtained through both the ML and ULS estimation methods are presented in Table 17.

![Parallel analysis graphical output](Figure 14. Parallel analysis graphical output)
Table 17. Original factor pattern

<table>
<thead>
<tr>
<th>Item</th>
<th>Loading on factor 1</th>
<th>Item</th>
<th>Loading on factor 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>A27_score</td>
<td>0.65306</td>
<td>A27_score</td>
<td>0.64939</td>
</tr>
<tr>
<td>A14_score</td>
<td>0.61638</td>
<td>A14_score</td>
<td>0.61856</td>
</tr>
<tr>
<td>A21_score</td>
<td>0.57646</td>
<td>A21_score</td>
<td>0.57375</td>
</tr>
<tr>
<td>A24_score</td>
<td>0.55926</td>
<td>A24_score</td>
<td>0.55911</td>
</tr>
<tr>
<td>A26_score</td>
<td>0.55832</td>
<td>A26_score</td>
<td>0.55220</td>
</tr>
<tr>
<td>A18_score</td>
<td>0.54786</td>
<td>A18_score</td>
<td>0.54422</td>
</tr>
<tr>
<td>A16_score</td>
<td>0.52559</td>
<td>A16_score</td>
<td>0.52691</td>
</tr>
<tr>
<td>A12_score</td>
<td>0.52279</td>
<td>A12_score</td>
<td>0.51902</td>
</tr>
<tr>
<td>A30_score</td>
<td>0.51861</td>
<td>A30_score</td>
<td>0.51506</td>
</tr>
<tr>
<td>A23_score</td>
<td>0.50278</td>
<td>A23_score</td>
<td>0.50318</td>
</tr>
<tr>
<td>A19_score</td>
<td>0.48722</td>
<td>A19_score</td>
<td>0.48427</td>
</tr>
<tr>
<td>A10_score</td>
<td>0.48212</td>
<td>A10_score</td>
<td>0.48404</td>
</tr>
<tr>
<td>A4_score</td>
<td>0.47677</td>
<td>A4_score</td>
<td>0.47833</td>
</tr>
<tr>
<td>A15_score</td>
<td>0.47502</td>
<td>A4_score</td>
<td>0.47810</td>
</tr>
<tr>
<td>A29_score</td>
<td>0.47106</td>
<td>A29_score</td>
<td>0.47180</td>
</tr>
<tr>
<td>A8_score</td>
<td>0.46221</td>
<td>A8_score</td>
<td>0.46559</td>
</tr>
<tr>
<td>A25_score</td>
<td>0.46218</td>
<td>A25_score</td>
<td>0.46214</td>
</tr>
<tr>
<td>A3_score</td>
<td>0.44598</td>
<td>A3_score</td>
<td>0.44888</td>
</tr>
<tr>
<td>A28_score</td>
<td>0.44279</td>
<td>A28_score</td>
<td>0.44321</td>
</tr>
<tr>
<td>A7_score</td>
<td>0.42239</td>
<td>A7_score</td>
<td>0.42426</td>
</tr>
<tr>
<td>A17_score</td>
<td>0.42233</td>
<td>A17_score</td>
<td>0.42232</td>
</tr>
<tr>
<td>A2_score</td>
<td>0.41204</td>
<td>A2_score</td>
<td>0.41467</td>
</tr>
<tr>
<td>A22_score</td>
<td>0.40420</td>
<td>A6_score</td>
<td>0.40321</td>
</tr>
<tr>
<td>A9_score</td>
<td>0.40152</td>
<td>A22_score</td>
<td>0.40237</td>
</tr>
<tr>
<td>A6_score</td>
<td>0.39647</td>
<td>A9_score</td>
<td>0.40179</td>
</tr>
<tr>
<td>A1_score</td>
<td>0.34938</td>
<td>A1_score</td>
<td>0.35086</td>
</tr>
<tr>
<td>A5_score</td>
<td>0.30577</td>
<td>A5_score</td>
<td>0.30926</td>
</tr>
<tr>
<td>A13_score</td>
<td>0.29389</td>
<td>A13_score</td>
<td>0.29422</td>
</tr>
<tr>
<td>A20_score</td>
<td>0.28599</td>
<td>A20_score</td>
<td>0.28895</td>
</tr>
<tr>
<td>A11_score</td>
<td>0.28555</td>
<td>A11_score</td>
<td>0.28647</td>
</tr>
</tbody>
</table>
To explore specific items within the MRQ, the loadings for each item were then examined, with a view to identifying any item with a loading less than .32 on the single factor retained. In reviewing individual item loadings within EFA, Pasta and Suhr (2004) recommend that: “For good interpretability it is helpful if (i) each factor has at least 3 items with high loadings (> 0.30); (ii) the variables that load on a factor share some conceptual meaning; (iii) the variables that load on different factors seem to measure different constructs; and (iv) the rotated factor pattern demonstrates simple structure – ideally each variable has a high loading on one factor and low loadings on other factors” (p. 4). In this case, the slightly more stringent requirement proposed by Tabachnick and Fidell (1992) of a minimum factor loading of .32 was adopted (i.e., equating to approximately 10% of overlapping variance between that item with other items loading on the single MRQ factor).

Based on the above analysis, MRQ items A5, A11, A13 and A20 were removed, and a second EFA conducted using both estimation methods. Table 18 presents the factor loadings from the second EFA which also yielded only one factor – which now accounted respectively for 82.19% and 81.68% of the total score variance, using ML and ULS estimation methods. All factor loadings in this second EFA exceeded .32, with the minimum loading observed to be .35.

Based on the EFA, it was concluded that the MRQ assesses one underlying construct, with 26 strongly loaded MRQ items. As highlighted above, since both ML and ULS estimation methods yielded similar results, it was concluded the slight non-normality of the data did not impact the EFA.
Table 18. Factor loadings with items removed

<table>
<thead>
<tr>
<th>Item</th>
<th>Loading on factor</th>
<th>Item</th>
<th>Loading on factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>A27_score</td>
<td>0.65828</td>
<td>A27_score</td>
<td>0.65555</td>
</tr>
<tr>
<td>A14_score</td>
<td>0.61441</td>
<td>A14_score</td>
<td>0.61700</td>
</tr>
<tr>
<td>A21_score</td>
<td>0.57614</td>
<td>A21_score</td>
<td>0.57344</td>
</tr>
<tr>
<td>A26_score</td>
<td>0.56450</td>
<td>A26_score</td>
<td>0.55933</td>
</tr>
<tr>
<td>A24_score</td>
<td>0.55848</td>
<td>A24_score</td>
<td>0.55816</td>
</tr>
<tr>
<td>A18_score</td>
<td>0.55351</td>
<td>A18_score</td>
<td>0.55094</td>
</tr>
<tr>
<td>A16_score</td>
<td>0.52406</td>
<td>A16_score</td>
<td>0.52570</td>
</tr>
<tr>
<td>A30_score</td>
<td>0.52399</td>
<td>A30_score</td>
<td>0.52140</td>
</tr>
<tr>
<td>A12_score</td>
<td>0.51746</td>
<td>A12_score</td>
<td>0.51228</td>
</tr>
<tr>
<td>A23_score</td>
<td>0.50437</td>
<td>A23_score</td>
<td>0.50538</td>
</tr>
<tr>
<td>A19_score</td>
<td>0.49210</td>
<td>A19_score</td>
<td>0.49029</td>
</tr>
<tr>
<td>A15_score</td>
<td>0.47729</td>
<td>A15_score</td>
<td>0.48205</td>
</tr>
<tr>
<td>A29_score</td>
<td>0.47644</td>
<td>A29_score</td>
<td>0.47881</td>
</tr>
<tr>
<td>A10_score</td>
<td>0.47641</td>
<td>A10_score</td>
<td>0.47723</td>
</tr>
<tr>
<td>A4_score</td>
<td>0.47540</td>
<td>A4_score</td>
<td>0.47671</td>
</tr>
<tr>
<td>A25_score</td>
<td>0.45869</td>
<td>A25_score</td>
<td>0.45766</td>
</tr>
<tr>
<td>A8_score</td>
<td>0.45352</td>
<td>A8_score</td>
<td>0.45563</td>
</tr>
<tr>
<td>A28_score</td>
<td>0.44563</td>
<td>A3_score</td>
<td>0.44842</td>
</tr>
<tr>
<td>A3_score</td>
<td>0.44528</td>
<td>A28_score</td>
<td>0.44656</td>
</tr>
<tr>
<td>A17_score</td>
<td>0.42846</td>
<td>A17_score</td>
<td>0.43044</td>
</tr>
<tr>
<td>A7_score</td>
<td>0.41866</td>
<td>A7_score</td>
<td>0.42017</td>
</tr>
<tr>
<td>A2_score</td>
<td>0.40766</td>
<td>A2_score</td>
<td>0.40989</td>
</tr>
<tr>
<td>A22_score</td>
<td>0.39791</td>
<td>A9_score</td>
<td>0.39588</td>
</tr>
<tr>
<td>A9_score</td>
<td>0.39659</td>
<td>A6_score</td>
<td>0.39519</td>
</tr>
<tr>
<td>A6_score</td>
<td>0.38961</td>
<td>A22_score</td>
<td>0.39458</td>
</tr>
<tr>
<td>A1_score</td>
<td>0.34928</td>
<td>A1_score</td>
<td>0.35106</td>
</tr>
</tbody>
</table>

Further evidence of internal structure of the MRQ was provided by the internal consistency data on the 26-item MRQ, using Cronbach’s α and item-rest correlations. Results indicated that the 26-item MRQ had a high level of internal consistency (Cronbach’s α=.89, item-rest correlations ranging from .33 to .61). These results, together with those from the EFA, provided strong support for the unidimensional nature of the MRQ measure.
4.2. CONFIRMATORY FACTOR ANALYSIS

While EFA is exploratory by nature, confirmatory factor analysis (CFA) is commonly used in multivariate statistical methods for testing the fit of specified theoretical models to empirical datasets was used. In this case, a CFA was used to confirm the results obtained through the EFA on the even-numbered random split-half sample. As above, the hypothetical model tested through the CFA was that the MRQ measured a unidimensional latent trait (moral reasoning). As Jackson, Gillaspy and Purc-Stephenson (2009) states:

CFA is often the analytic tool of choice for developing and refining measurement instruments, assessing construct validity, identifying method effects, and evaluating factor invariance across time and groups (Brown, 2006) (p.6).

As with EFA, SAS (version 9.4) was used to perform CFA. Preliminary analyses of the data indicated no missing cases, with univariate and multivariate tests for normality on the even-numbered observations (both for the MRQ total score and for the individual MRQ item scores) indicating skewness and kurtosis coefficients that fell well within accepted thresholds (skewness < 2 and kurtosis < 7 – see West, Finch and Curran, 1995).

Various authors have noted that while CFA assumes multivariate normality, the ML method is generally robust to slight departures from normality (West, Finch & Curran, 1995). Nonetheless, the ML method may be unsuitable for severe violations of normality as the chi-square statistic may be over-estimated (Jackson,
Gillaspy & Purc-Stephenson, 2009). To reduce any risk of bias in the chi-square statistics obtained here, the ML with Satorra-Bentler scaled chi-square statistics for model fit (MLSB) was used. The MLSB is also sometimes labelled Robust ML (RML), generally producing less biased standard errors in cases where data are not strictly normal. The appropriateness of MLSB for non-normal data in a CFA is summarised by Motl, Dishman, Birnbaum and Lytle (2005):

The Satorra-Bentler procedure is a normal theory method that adjusts the chi-square statistic and standard errors for the non-normality of the data, and the correction yields more accurate goodness-of-fit statistics and standard errors than do other methods developed for nonnormal data (Chou, Bentler, & Satorra, 1991; Fouladi, 2000; Hu, Bentler, & Kano, 1992) (p.94).

For the purposes of further triangulation, comparisons of the results obtained with those using the ULS method were also conducted, given that ULS does not require any distributional assumptions (Jöreskog, 2003). In all, four CFAs were performed on the even-numbered case set, to ensure that the results were robust across a range of different conditions. First, a CFA was performed using the MLSB method on the covariance matrix from the full original set of even-numbered data. Second, the MLSB method was used to perform a CFA on the MRQ, excluding items A5, A11, A13 and A20 (as suggested by the EFA results). Both of these steps were then repeated using the ULS method, to confirm that the modest non-normality observed in the item distributions did not have any
substantial impact on the results obtained. Table 19 presents the goodness-of-fit indicators for the four estimation methods.

Table 19. Goodness-of-fit indicators of estimation methods

<table>
<thead>
<tr>
<th>Estimation method</th>
<th>$\chi^2$</th>
<th>$\chi^2$/df</th>
<th>df</th>
<th>CFI</th>
<th>NFI</th>
<th>RMSEA</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLSB</td>
<td>629.97*</td>
<td>-</td>
<td>405</td>
<td>1.56</td>
<td>0.90</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>MLSB with items A5, A11, A13, A20 removed</td>
<td>459.62*</td>
<td>170.35</td>
<td>299</td>
<td>1.53</td>
<td>0.93</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>ULS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.96</td>
<td>-</td>
</tr>
<tr>
<td>ULS with items A5, A11, A13, A20 removed</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.97</td>
<td>-</td>
</tr>
</tbody>
</table>

* $p < .001$

The interpretation of these indices was guided by recommended thresholds of commonly reported goodness-of-fit indices. The cut-off criteria for fit indices recommended by Hu and Bentler (1999), and those drawn from a review of CFA reporting practices by Schreiber, Stage, King, Nora and Barlow (2006), as well as recommendations by Hair, Black, Babin and Anderson (2009) and by Brown (2014) were all considered in the present study. Based on a recommendation by Brown (2014), at least one fit index from each of the following three categories should be reported in any presentation of CFA results: (i) absolute fit indices (e.g., $\chi^2$ and the SRMR); (ii) parsimony correction indices (e.g., the RMSEA); and (iii) comparative or incremental fit indices (e.g., the CFI and NFI). The results obtained are discussed in terms of each of these types of indices below.

Based on the absolute fit indices reported in Table 19, while the $\chi^2$ values obtained were significant, the $\chi^2$/df values were less than 2 in both MLSB estimation cases.
This indicates a good overall model fit to the data despite the significant $\chi^2$ obtained, given the sensitivity of the $\chi^2$ statistic to sample size variations (Brown, 2014; Hair et al., 2009). The SRMR, which indicates the average discrepancy between correlations drawn from the data and those of the predicted model, has a range of 0.0 to 1.0, with the former indicating a perfect model fit. In this case, SRMR values of all MLSB and ULS estimation cases were less than .05. This indicates a good model fit, given that: (i) Hu and Bentler (1999) suggested a cut-off value for the SRMR of .08; and (ii) Hair et al.’s (2009) recommendation that the SRMR should not exceed 0.1.

The RMSEA was used here as the parsimony correction index, to provide an index of absolute fit or badness-of-fit of the model to the data (Hair et al., 2009). The RSMEA ranges 0.0 to 1.0 with the former indicating perfect model fit. Hu and Bentler (1999) recommended that, as a rule of thumb, the RMSEA value should ideally be less than or equal to .05. As the RMSEA values of both MLSB estimation cases were .04, falling between .03 and .04 with a 90% confidence limit, these results suggest a good fit between the theoretical structure of MRQ and the empirical results obtained.

Comparative or incremental fit indices evaluate the fit of a user-specified solution against an alternative restricted baseline model (Brown, 2014; Hair et al., 2009). The CFI, a widely used index of comparative fit, ranges 0.0 to 1.0, with the latter indicating perfect model fit. In this case, the CFI obtained for the MLSB estimation with items removed was .93, and was higher than that for the MLSB estimate with all items included. Thus, the MLSB estimation with items removed
showed a good overall model fit, that was superior to that using the full item sample, given that (i) The CFI was above the .90 threshold suggested by Hair et al.; and that (ii) the CFI was close to the .95 acceptance threshold for a relatively good model fit between the hypothesised model and observed data suggested by Hu and Bentler (1999). While the NFI is used less today, this index was also used here to provide information on the ULS estimation results. The NFI ranges 0.0 to 1.0 with the latter indicating a perfect fit (Hair et al., 2009). In this case, the NFI value both for the ULS estimation exceeded .96, which suggests a good model fit. Further, the NFI for the ULS estimation with items A5, A11, A13 and A20 removed was .97. This suggests again that a better model fit was achieved with these four items removed.

Collectively, the fit indices from the MLSB estimation with items A5, A11, A13 and A20 removed indicated a good overall model fit to the data (see Figure 15), $\chi^2 (299, N = 335) = 459.62, p < .001, SRMR = .05, RMSEA = .04, CFI = .93$. Further, the slight non-normality of the data did not appear to have impacted the analysis outcomes, given that similar results were obtained through the use of the ULS estimation method ($NFI = .97, SRMR = .05$).
The standardised loading estimates on the latent trait (moral reasoning) and the standardised residuals of item pairs also supported the notion that the results indicated sound model fit to the data. All loadings obtained ranged from .42 to .66, and were statistically significant \( (p < .001) \), suggesting non-chance relationships between the items and the overall trait being assessed (moral reasoning). Of the 26 items analysed, three items (A2, 9 and 22) had loading estimates ranging .42 to .44; this range is slightly less than the .5 recommended threshold of standardised loading estimate by Hair et al. (2009). Standardised residuals provide information to identify item pairs for which the a priori measurement model does not accurately predict observed covariances between a pair of items. In this case, all standardised residuals of item pairs were less than 4.0, with the exception of the item pair A15-A29 (4.49). These results would generally be considered acceptable, as recommended by Hair et al. (2009).

The average variance extracted (AVE) and construct reliability (CR) coefficients were .3 and .9 respectively. While an AVE value of less than .5 suggests the possibility of more error variance in the items than variance explained by the latent factor on average, the CR coefficient exceeds the recommended .7. These suggest that the moral reasoning measurement model has adequate convergence and construct reliability (Hair et al., 2007).
Figure 15. Measurement model with MLSB estimation

*: p < .001
4.3. Rasch Analysis of Item Scores

A Rasch analysis was conducted using the Rasch Unidimensional Measurement Model (RUMM2030) software to complement the EFA and CFA findings. The MRQ items were assessed, using parametric statistical tests, for: (i) threshold ordering and internal reliability; (ii) the overall model fit using the overall fit statistics; (iii) individual item and person fit; (iv) the item characteristic curves; (v) local dependency and differential item functioning; and (vi) dimensionality.

Despite its widespread application, CTT is subject to several major limitations, owing in large part to its circular dependency in terms of conceptualisation. In CTT, person statistics (i.e., observed scores) are inherently item-sample dependent, while item statistics (i.e., difficulty and discrimination) are person-sample dependent. This restricts the applicability of CTT in various important measurement situations (e.g., situations in which different tests must be equated). Further, in situations that involve rating scale data, approaches grounded in CTT do not provide a basis for exploring additivity of scores, a critical attribute of any valid measure. Rasch modelling offers an alternative basis for constructing measurements within education. Rasch models incorporate the ordering features of Guttman scaling, and focus on probabilistic distributions of examinees’ performance at the item level, rather than on test-level information, like CTT. An important advantage of the Rasch approach derives from the fact that the item parameters obtained do not depend on the characteristics of the persons taking the test, and that the person parameters do not depend on the specific items chosen for a given test. The parameters produced through Rasch
analysis are thus independent of specific sample characteristics. This approach addresses many of the weaknesses of the CTT approach. Rasch measurement models are also tenable as a basis for examining scores obtained through rating scales, because these models provide a means by which the hierarchical structure, unidimensionality and additivity of the scores can be evaluated.

Noting the polytomous scoring structure of the MRQ, either a rating scale or partial credit parameterisation could be applied to the entire set of responses for the analysis. The likelihood ratio test that assesses the efficiency of parameterisation used for the unrestricted form of the Rasch model on the entire dataset was significant and hence indicated: (i) inconsistencies in the distance between response thresholds; and (ii) whether the unrestricted model contains more information than the rating model. Andrich, Lyne, Sheridan and Luo (2015), however, stated that:

The Rasch model will provide a set of item threshold estimates for each item based on the specifications provided for each individual item. The use of terms such as Dichotomous Model, Partial Credit Model, Rating Model, is therefore not only unnecessary but provides a misleading expectation that there are several, distinct and different Rasch models when, if fact, there is only the ONE model involved with all these different types of items and all with the same response structure (p. 60).
4.3.1. Threshold Ordering and Reliability

In this study, the Rasch analysis was performed primarily to assess the data fit to an unrestricted Rasch model, without assuming a uniform distance between response thresholds. The initial results obtained from the Rasch analysis indicated the presence of disordered thresholds across a number of items in the MRQ, and a significant chi-square statistic, \( \chi^2 (270, N = 669) = 517.27, p < .001 \). The reliability of the MRQ, however, appeared to be good, with a person-separation index (PSI) of .88. This high PSI indicated: (i) a good spread of item estimates given the presence of multiple thresholds (six response categories) for each item, and (ii) a high estimated true variance in respondents’ moral reasoning levels (i.e., only 12% of the variance attributable to error variance).

The presence of disordered thresholds suggested that respondents might not have been able to distinguish between the six rank order options presented within the MRQ based on the Rasch model. Given this result, the categories were collapsed into a smaller number, and the collapsed data subjected to a second round of analysis. The collapsed categories were premised on the following assumptions: (i) that a respondent would score 2 if s/he identified the pre-conventional level as the lowest level of moral judgment; (ii) a respondent would score 0 if s/he identified the levels of moral judgment opposite to that of Kohlberg’s framework; and (iii) a respondent would score 1 for all other rank order permutations. Table 20 presents the updated scoring matrix with collapsed categories.
Table 20. Scoring matrix of two-tier items with collapsed categories

<table>
<thead>
<tr>
<th>Rank order (levels of moral judgment)</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Post-conventional (2) Conventional (3) Pre-conventional</td>
<td>2</td>
</tr>
<tr>
<td>(1) Conventional (2) Post-conventional (3) Pre-conventional</td>
<td>2</td>
</tr>
<tr>
<td>(1) Post-conventional (2) Pre-conventional (3) Conventional</td>
<td>1</td>
</tr>
<tr>
<td>(1) Conventional (2) Pre-conventional (3) Post-conventional</td>
<td>1</td>
</tr>
<tr>
<td>(1) Pre-conventional (2) Post-conventional (3) Conventional</td>
<td>1</td>
</tr>
<tr>
<td>(1) Pre-conventional (2) Conventional (3) Post-conventional</td>
<td>0</td>
</tr>
</tbody>
</table>

A second Rasch analysis was then performed on the MRQ using the collapsed categories. Items A5, A11, A13 and A20 were removed from this analysis, as disordered thresholds remained evident in these items despite collapsing categories. All other items did not have disordered thresholds, using the collapsed categories. This is in agreement with the CFA that also identified these four items as items that likely caused a model misfit. Figure 16 presents the threshold map for all 26 items (with A5, A11, A13 and A20 removed) and Table 21 shows the uncentralised item thresholds that indicate all category responses have been used as expected consistently.
<table>
<thead>
<tr>
<th>Item label in RUMM2030</th>
<th>Item number in MRQ</th>
<th>Endorsability</th>
<th>Mean</th>
<th>Threshold 1</th>
<th>Threshold 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>I0001</td>
<td>A1</td>
<td>.01</td>
<td>.01</td>
<td>-.60</td>
<td>.63</td>
</tr>
<tr>
<td>I0002</td>
<td>A2</td>
<td>.04</td>
<td>.04</td>
<td>-.37</td>
<td>.45</td>
</tr>
<tr>
<td>I0003</td>
<td>A3</td>
<td>-.01</td>
<td>-.01</td>
<td>-.46</td>
<td>.44</td>
</tr>
<tr>
<td>I0004</td>
<td>A4</td>
<td>-.08</td>
<td>-.08</td>
<td>-.86</td>
<td>.69</td>
</tr>
<tr>
<td>I0005</td>
<td>A6</td>
<td>.17</td>
<td>.17</td>
<td>-.59</td>
<td>.94</td>
</tr>
<tr>
<td>I0006</td>
<td>A7</td>
<td>1.00</td>
<td>1.00</td>
<td>-.41</td>
<td>2.41</td>
</tr>
<tr>
<td>I0007</td>
<td>A8</td>
<td>.33</td>
<td>.33</td>
<td>-1.08</td>
<td>1.74</td>
</tr>
<tr>
<td>I0008</td>
<td>A9</td>
<td>-.45</td>
<td>-.45</td>
<td>-2.86</td>
<td>1.95</td>
</tr>
<tr>
<td>I0009</td>
<td>A10</td>
<td>.11</td>
<td>.11</td>
<td>-.70</td>
<td>.92</td>
</tr>
<tr>
<td>I0010</td>
<td>A12</td>
<td>-.06</td>
<td>-.06</td>
<td>-1.01</td>
<td>.88</td>
</tr>
<tr>
<td>I0011</td>
<td>A14</td>
<td>.09</td>
<td>.09</td>
<td>-.54</td>
<td>.72</td>
</tr>
<tr>
<td>I0012</td>
<td>A15</td>
<td>-.78</td>
<td>-.78</td>
<td>-1.21</td>
<td>-.35</td>
</tr>
<tr>
<td>I0013</td>
<td>A16</td>
<td>-.51</td>
<td>-.51</td>
<td>-1.47</td>
<td>.45</td>
</tr>
<tr>
<td>I0014</td>
<td>A17</td>
<td>-.24</td>
<td>-.24</td>
<td>-1.21</td>
<td>.73</td>
</tr>
<tr>
<td>I0015</td>
<td>A18</td>
<td>-.00</td>
<td>-.00</td>
<td>-1.20</td>
<td>1.20</td>
</tr>
<tr>
<td>I0016</td>
<td>A19</td>
<td>-.66</td>
<td>-.66</td>
<td>-1.09</td>
<td>-.22</td>
</tr>
<tr>
<td>I0017</td>
<td>A21</td>
<td>.33</td>
<td>.33</td>
<td>-.47</td>
<td>1.12</td>
</tr>
<tr>
<td>I0018</td>
<td>A22</td>
<td>-.46</td>
<td>-.46</td>
<td>-1.53</td>
<td>.62</td>
</tr>
<tr>
<td>I0019</td>
<td>A23</td>
<td>.93</td>
<td>.93</td>
<td>-.05</td>
<td>1.91</td>
</tr>
<tr>
<td>I0020</td>
<td>A24</td>
<td>-.17</td>
<td>-.17</td>
<td>-.68</td>
<td>.33</td>
</tr>
<tr>
<td>I0021</td>
<td>A25</td>
<td>.19</td>
<td>.19</td>
<td>-.38</td>
<td>.77</td>
</tr>
<tr>
<td>I0022</td>
<td>A26</td>
<td>.18</td>
<td>.18</td>
<td>-.52</td>
<td>.88</td>
</tr>
<tr>
<td>I0023</td>
<td>A27</td>
<td>.59</td>
<td>.59</td>
<td>-.34</td>
<td>1.51</td>
</tr>
<tr>
<td>I0024</td>
<td>A28</td>
<td>.16</td>
<td>.16</td>
<td>-.77</td>
<td>1.08</td>
</tr>
<tr>
<td>I0025</td>
<td>A29</td>
<td>-.48</td>
<td>-.48</td>
<td>-1.05</td>
<td>.09</td>
</tr>
<tr>
<td>I0026</td>
<td>A30</td>
<td>-.21</td>
<td>-.21</td>
<td>-.85</td>
<td>.43</td>
</tr>
</tbody>
</table>
4.3.2. Overall Model Fit

In terms of the overall fit of the model, while the $\chi^2$ test for the overall fit to the Rasch model was significant at $\chi^2 (234, n = 647) = 333.74, p < .001$, it could have been due to its sensitivity to the large sample size. With an adjusted sample, the overall fit test indicated $\chi^2 (234, n = 527) = 271.84, p = .05$. This non-significant item-trait interaction chi-square statistic suggests: (i) a good overall model fit; and (ii) that the items collectively measure a common latent trait. The good overall model fit was also supported by the item fit residual ($M = -.14, SD = 1.53$) and person fit residual ($M = -.16, SD = .95$), which had means close to 0 and SDs close to 1, as suggested by Tennant and Conaghan (2007).

With respect to reliability, the PSI of .84 and Cronbach’s $\alpha$ of .89 suggests acceptable reliability as the minimum $\text{PSI} \geq .7$ (Tennant & Conaghan, 2007). Hence, these outcomes suggest that the MRQ scale should be able to differentiate between at least two groups of respondents.

4.3.3. Individual Item and Person Fit

Examining the individual item and person fit outputs, fit residuals should lie within the range of ± 2.5 for an item to be fitting to the Rasch model (Tennant & Conaghan, 2007; Tennant & Pallant, 2006). In this case, all items (with Bonferroni adjustment) except for A9 (2.71) and 26 (-2.65) were within the threshold and $p$-values were > .01 (see Table 22). Based on this evidence, the vast majority of items within the MRQ met criteria for adequate fit.
4.3.4. Item Characteristic Curves

The item characteristic curve (ICC) of A9 (Figure 17) suggests that respondents at the post-conventional level of moral reasoning found it difficult to endorse, which influenced the item fit. The ICC for A26 (Figure 18) suggests that the item misfit observed could have been influenced by over-discrimination where respondents at the post-conventional level found the item too endorsable while those at the pre-conventional level found it too difficult to endorse.

Though the fit residuals of A9 and A26 were slightly beyond the ± 2.5 range, their $\chi^2$ probability with Bonferroni adjustment was not < .01. Hence, it was concluded that all 26 items measure a common underlying construct. For person fit, 10 respondents excluding extreme cases had a fit residual out of the ± 2.5 range between -3.69 and 3.11. This could indicate anomalies in the score patterns of these respondents, which may have reflected fatigue. As there was no data entry error and given the good overall model fit, these respondents were not removed.
<table>
<thead>
<tr>
<th>Item label in RUMM2030</th>
<th>Item number in MRQ</th>
<th>Fit residual</th>
<th>$\chi^2$ probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>I0001</td>
<td>A1</td>
<td>1.11</td>
<td>0.22</td>
</tr>
<tr>
<td>I0002</td>
<td>A2</td>
<td>0.73</td>
<td>0.60</td>
</tr>
<tr>
<td>I0003</td>
<td>A3</td>
<td>-0.47</td>
<td>0.78</td>
</tr>
<tr>
<td>I0004</td>
<td>A4</td>
<td>-0.19</td>
<td>0.84</td>
</tr>
<tr>
<td>I0005</td>
<td>A6</td>
<td>1.68</td>
<td>0.10</td>
</tr>
<tr>
<td>I0006</td>
<td>A7</td>
<td>0.32</td>
<td>0.14</td>
</tr>
<tr>
<td>I0007</td>
<td>A8</td>
<td>1.27</td>
<td>0.31</td>
</tr>
<tr>
<td>I0008</td>
<td>A9</td>
<td>2.71</td>
<td>0.01</td>
</tr>
<tr>
<td>I0009</td>
<td>A10</td>
<td>0.12</td>
<td>0.11</td>
</tr>
<tr>
<td>I0010</td>
<td>A12</td>
<td>1.70</td>
<td>0.01</td>
</tr>
<tr>
<td>I0011</td>
<td>A14</td>
<td>-2.35</td>
<td>0.06</td>
</tr>
<tr>
<td>I0012</td>
<td>A15</td>
<td>-0.99</td>
<td>0.19</td>
</tr>
<tr>
<td>I0013</td>
<td>A16</td>
<td>-2.02</td>
<td>0.03</td>
</tr>
<tr>
<td>I0014</td>
<td>A17</td>
<td>0.85</td>
<td>0.88</td>
</tr>
<tr>
<td>I0015</td>
<td>A18</td>
<td>-2.25</td>
<td>0.02</td>
</tr>
<tr>
<td>I0016</td>
<td>A19</td>
<td>-1.41</td>
<td>0.36</td>
</tr>
<tr>
<td>I0017</td>
<td>A21</td>
<td>-2.23</td>
<td>0.07</td>
</tr>
<tr>
<td>I0018</td>
<td>A22</td>
<td>0.63</td>
<td>0.77</td>
</tr>
<tr>
<td>I0019</td>
<td>A23</td>
<td>-0.02</td>
<td>0.52</td>
</tr>
<tr>
<td>I0020</td>
<td>A24</td>
<td>-1.39</td>
<td>0.29</td>
</tr>
<tr>
<td>I0021</td>
<td>A25</td>
<td>1.16</td>
<td>0.20</td>
</tr>
<tr>
<td>I0022</td>
<td>A26</td>
<td>-2.65</td>
<td>0.01</td>
</tr>
<tr>
<td>I0023</td>
<td>A27</td>
<td>-2.08</td>
<td>0.06</td>
</tr>
<tr>
<td>I0024</td>
<td>A28</td>
<td>1.32</td>
<td>0.18</td>
</tr>
<tr>
<td>I0025</td>
<td>A29</td>
<td>1.24</td>
<td>0.29</td>
</tr>
<tr>
<td>I0026</td>
<td>A30</td>
<td>-0.46</td>
<td>0.60</td>
</tr>
</tbody>
</table>
4.3.5. Local Dependency and Differential Item Functioning

With respect to local dependency, from the residual correlation matrix, the maximum inter-item residual correlation ($r = .18$) was between items A2 and A3, less than the .2 threshold recommended by Andrich, Lyne, Sheridan and Luo (2015). This indicates minimal local dependency and how a respondent performs for an item will have little or no bearing on other items.
Differential item functioning (DIF) analyses were also performed to determine whether there was any evidence of item bias within the MRQ, although these analyses were somewhat limited by the data that could be obtained in the study. As the MRQ is designed to assess moral reasoning across students at secondary one to four across gender, schools and streams, DIF was assessed for school, stream, gender and level. A Bonferroni correction (dividing the probability value of significance by the number of tests of fit) was also used in these analyses.

Table 23 presents the p-values used to determine the existence of DIF by level, gender, school and educational stream. From the results, it can be concluded that no DIF was evident for school, educational stream, gender and level, with the exception of item A10 (p < .001) which exhibited item bias across different levels.
Table 23. DIF by level, gender, school and education stream

<table>
<thead>
<tr>
<th>Item number</th>
<th>p-value</th>
<th>by level</th>
<th>by gender</th>
<th>by school</th>
<th>by educational stream</th>
</tr>
</thead>
<tbody>
<tr>
<td>I0001</td>
<td>0.033</td>
<td>0.508</td>
<td>0.233</td>
<td>0.516</td>
<td></td>
</tr>
<tr>
<td>I0002</td>
<td>0.715</td>
<td>0.621</td>
<td>0.893</td>
<td>0.085</td>
<td></td>
</tr>
<tr>
<td>I0003</td>
<td>0.083</td>
<td>0.146</td>
<td>0.121</td>
<td>0.368</td>
<td></td>
</tr>
<tr>
<td>I0004</td>
<td>0.530</td>
<td>0.006</td>
<td>0.060</td>
<td>0.247</td>
<td></td>
</tr>
<tr>
<td>I0005</td>
<td>0.597</td>
<td>0.676</td>
<td>0.736</td>
<td>0.545</td>
<td></td>
</tr>
<tr>
<td>I0006</td>
<td>0.552</td>
<td>0.007</td>
<td>0.043</td>
<td>0.377</td>
<td></td>
</tr>
<tr>
<td>I0007</td>
<td>0.181</td>
<td>0.100</td>
<td>0.346</td>
<td>0.572</td>
<td></td>
</tr>
<tr>
<td>I0008</td>
<td>0.057</td>
<td>0.428</td>
<td>0.086</td>
<td>0.343</td>
<td></td>
</tr>
<tr>
<td>I0009</td>
<td>0.000</td>
<td>0.475</td>
<td>0.344</td>
<td>0.927</td>
<td></td>
</tr>
<tr>
<td>I0010</td>
<td>0.009</td>
<td>0.136</td>
<td>0.078</td>
<td>0.089</td>
<td></td>
</tr>
<tr>
<td>I0011</td>
<td>0.472</td>
<td>0.731</td>
<td>0.658</td>
<td>0.580</td>
<td></td>
</tr>
<tr>
<td>I0012</td>
<td>0.247</td>
<td>0.208</td>
<td>0.270</td>
<td>0.939</td>
<td></td>
</tr>
<tr>
<td>I0013</td>
<td>0.602</td>
<td>0.862</td>
<td>0.482</td>
<td>0.258</td>
<td></td>
</tr>
<tr>
<td>I0014</td>
<td>0.281</td>
<td>0.005</td>
<td>0.262</td>
<td>0.224</td>
<td></td>
</tr>
<tr>
<td>I0015</td>
<td>0.441</td>
<td>0.063</td>
<td>0.010</td>
<td>0.266</td>
<td></td>
</tr>
<tr>
<td>I0016</td>
<td>0.884</td>
<td>0.878</td>
<td>0.269</td>
<td>0.227</td>
<td></td>
</tr>
<tr>
<td>I0017</td>
<td>0.897</td>
<td>0.954</td>
<td>0.342</td>
<td>0.600</td>
<td></td>
</tr>
<tr>
<td>I0018</td>
<td>0.308</td>
<td>0.262</td>
<td>0.099</td>
<td>0.220</td>
<td></td>
</tr>
<tr>
<td>I0019</td>
<td>0.748</td>
<td>0.049</td>
<td>0.452</td>
<td>0.038</td>
<td></td>
</tr>
<tr>
<td>I0020</td>
<td>0.514</td>
<td>0.682</td>
<td>0.577</td>
<td>0.325</td>
<td></td>
</tr>
<tr>
<td>I0021</td>
<td>0.318</td>
<td>0.816</td>
<td>0.047</td>
<td>0.637</td>
<td></td>
</tr>
<tr>
<td>I0022</td>
<td>0.606</td>
<td>0.688</td>
<td>0.967</td>
<td>0.032</td>
<td></td>
</tr>
<tr>
<td>I0023</td>
<td>0.305</td>
<td>0.013</td>
<td>0.187</td>
<td>0.284</td>
<td></td>
</tr>
<tr>
<td>I0024</td>
<td>0.421</td>
<td>0.853</td>
<td>0.441</td>
<td>0.827</td>
<td></td>
</tr>
<tr>
<td>I0025</td>
<td>0.954</td>
<td>0.115</td>
<td>0.130</td>
<td>0.307</td>
<td></td>
</tr>
<tr>
<td>I0026</td>
<td>0.269</td>
<td>0.475</td>
<td>0.149</td>
<td>0.088</td>
<td></td>
</tr>
</tbody>
</table>
Figure 19 presents the item characteristic curve with the level plots for A10 of the MRQ. Visually, there is item bias especially between the secondary one (S1) and secondary four (S4) levels at 0 to 0.5 logits and above 2.5 logits. The DIF could be significant due to the large sample size given that a small difference in how an item functions across sub-groups would result in a significant statistical test. With item A10 removed, the model fit was not impacted. In view of this and the evidence supporting the adequacy of how the data fits the Rasch model, item A10 was eventually retained.

![Figure 19. ICC with the level plots for A10 (item 9 in RUMM2030)](image)

**4.3.6. Dimensionality**

Further to the non-significant item-trait interaction chi-square statistic discussed earlier, evidence from the principal components analysis (PCA) support the unidimensionality of the MRQ. All the items loaded only on one principal component, which supports the assumption of item local independence. As proposed by Smith (2002), the PCA can reveal two patterns of items that load
(positively and negatively) onto the first principal component. An independent $t$-test can be done on these two sets of items to make separate person estimates. For unidimensionality, no more than 5% of the $t$-test results should be significant ($p < .05$). As only 4.33% of respondents showed a significant difference between the person locations based on the two sets of items, the unidimensionality of the MRQ was supported (see Figure 20).

Figure 20. PCA t-test of +/- loaded items on first principal component

The findings above point to a good overall model fit based on the Rasch model, and support the unidimensionality of the MRQ. The person-item threshold distribution that places student (person) and item location estimates on the same logit scale (see Figure 21) shows that the items and thresholds spanned almost the range of person scores except for some who scored very high on the MRQ. This could be explained by the MOE’s expectation that more secondary school
students fall within the conventional to post-conventional levels of moral development. While this suggests that the items measure moral reasoning from the pre-conventional to post-conventional levels, future versions could include more ‘difficult’ items to assess the conventional to post-conventional levels.

Psychometric properties of the MRQ through the Rasch analysis have been adequate in that there is good fit of the data to the Rasch unidimensional model. Further, analyses suggested that inferences drawn from the measure should not be confounded by students’ demographic attributes (e.g., gender, school or stream). Females did have slightly higher moral reasoning scores, but did not differ significantly from males \( (F(1, 668) = 2.91, p = .09) \) (see Figure 22). Moral reasoning scores were also not influenced by the school \( (F(2, 667) = .09, p = .92) \) (see Figure 23) or by the stream \( (F(1, 668) = .31, p = .58) \) (see Figure 24) that the students were from. Nonetheless, for the levels of study, there was a statistical significant difference in moral reasoning scores \( (F(3, 666) = 2.87, p = .04) \) (see

![Figure 21. Person-item threshold distribution](image-url)
Figure 25). While this difference is to be expected, as moral reasoning should be developmental, there was no definitive trend in the analysis that students at higher levels of study scored higher on the MRQ.

Figure 22. Person-item measure threshold distribution by gender

Figure 23. Person-item measure threshold distribution by school
The Rasch analysis served to complement the EFA and CFA findings. Based on the discussion of the six areas (i.e., threshold ordering and internal reliability, overall model fit, individual item and person fit, item characteristic curves, local dependency and differential item functioning, and dimensionality), there was adequate evidence to affirm the unidimensionality of the MRQ. Items also
proved to be functioning as anticipated, based on how the data fit the Rasch model. This being the case, a linear person-measure of the MRQ can also be established for the meaningful comparison of respondents’ moral reasoning.

4.4. EFA ON COLLAPSED MRQ SCORES

Given that the MRQ response categories were collapsed after the Rasch analysis was performed, an EFA and parallel analysis was then re-conducted with the odd-numbered dataset to explore whether the original factor structure was retained after the scoring categories were collapsed for the 26 items. These analyses indicated that the MRQ scores based on collapsed categories still demonstrated high internal consistency, Cronbach’s $\alpha = .87$ (Kline, 2000). There was also minimal to no risk of item redundancy based on the recommended $\alpha$ threshold of .9 (Streiner, 2003). The Tucker and Lewis reliability coefficient was also computed as high (.91). Based on Kaiser’s (1958) eigenvalue rule and Cattell’s scree plot in Figure 26, the MRQ with collapsed categories yielded one factor that could explain 81.38% of variance for the entire set of variables.
A parallel analysis done with a Monte Carlo simulation of 1000 simulated datasets indicated one strong factor, similar to Cattel’s scree plot and Kaiser’s eigenvalue rule. The graphical output of the parallel analysis for the MRQ with collapsed categories is presented in Figure 27. Based on these results, the EFA extracted one factor without rotation. Factor loadings of ranged from an acceptable .31 to .61.

Figure 26. Scree plot and variance explained figures by ML estimation*

* Based on re-scored items
4.5. CFA ON COLLAPSED MRQ SCORES

Following the EFA, a CFA was then re-conducted on the even-numbered cases using the collapsed response categories. Table 24 presents results from the latter CFA. Evidently, the CFA of the MRQ with collapsed categories presented fit indices that support an overall good model fit to the hypothesised measurement model. Supporting the overall good model fit are the standardised loading estimates on the latent trait (moral reasoning) and the standardised residuals of item pairs. All loadings ranged from .38 to .64, a slight decrease from the initial CFA without collapsing categories that ranged between .42 and .66. Nonetheless, all standardised loading estimates were statistically significant indicating non-chance relationships with moral reasoning. Of the 26 items analysed, six items...
(A8, 9, 12, 17, 25 and 30) had loading estimates is slightly less than the recommended .5 threshold of standardised loading estimate by Hair et al. (2009).

Table 24. Goodness-of-fit indicators of CFA with collapsed categories

<table>
<thead>
<tr>
<th>Estimation method</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$\chi^2$/df</th>
<th>CFI</th>
<th>RMSEA</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLSB with items A5, A11, A13, A20 removed</td>
<td>441.56*</td>
<td>299</td>
<td>1.48</td>
<td>0.93</td>
<td>0.04</td>
<td>0.05</td>
</tr>
</tbody>
</table>

*p < .001

All standardised residuals of item pairs in the second CFA were less than 4.0 with the exception of the item pair A2-A3 (4.81) and A15-A19 (4.23). This is considered acceptable as recommended by Hair et al. (2009). Figure 28 presents the path diagram illustrating the standardised loading estimates for the CFA of the MRQ with collapsed categories. The average variance extracted (AVE) and construct reliability (CR) coefficients remained at .3 and .9 respectively. Again, while an AVE value of less than .5 suggests the possibility of more error variance in the items than variance explained by the latent factor on average, the CR coefficient exceeds the recommended .7. These results suggest that the rescored moral reasoning measurement model has adequate convergence and construct reliability (Hair et al., 2007).
Figure 28. Path diagram with standardised loading estimates
4.6. **Measurement Invariance Across Sub-Groups**

After satisfactory psychometric properties of the MRQ had been established, a multi-group CFA (MGCFA) was done using SAS 9.4 to ascertain the extent of measurement invariance associated with the MRQ. It should be noted here that an MGCFA does not seek to determine whether there is measurement equivalence across groups, but the degree of equivalence across groups. If there is lacking evidence supporting measurement invariance, interpretations between groups may not be meaningful (Vandenberg & Lance, 2000).

Pendergast, von der Embse, Kilguss and Eklund (2017) suggest that MGCFA is typically and notably used to evaluate measurement invariance across two groups. If the number of groups increase, MGCFA may be overly complex and other approaches such as Multiple Indicators Multiple Causes (MIMIC) modelling may be more appropriate. Further, MGCFA, being an extension of CFA, generally requires a minimum of 200 responses per group (Pendergast et al., 2017). This is supported in a study by Chen (2007) who concluded that inadequate sample sizes \((n < 300)\) led to less stable and less accurate goodness-of-fit indices.

The hierarchical levels in order of increasing restrictive constraints in MGCFA are: (i) configural invariance, where no constraints are placed across both models; (ii) metric (weak) invariance, where only factor loadings are constrained for both models; (iii) scalar (strong) invariance, where factor loadings and corresponding indicator means and intercepts are constrained for both models; and (iv) strict invariance, where factor loadings, means and intercepts, and residual error
variances are constrained and fitted with the same model (Yung, 2010; Schulz, 2016; Pendergast et al., 2017)

Measurement invariance can be considered established when at least two of the change in goodness-of-fit indices are within the recommended thresholds: (i) $\Delta \chi^2$ results in $p > .05$; (ii) $\Delta CFI < -.01$ (with definite differences between models if $\Delta CFI > -.02$; (iii) $\Delta TLI < -.01$; and (iv) $\Delta RMSEA < .015$ (Cheung & Rensvold, 2002; Chen, 2007). However, it is noteworthy that, as with CFA, $\Delta \chi^2$ can be sensitive to sample size and $\Delta CFI$, $\Delta RMSEA$ and $\Delta TLI$ are not influenced by sample size - it is not desirable to rely on $\Delta \chi^2$ to determine measurement invariance (Vandenberg & Lance, 2000; Cheung & Rensvold, 2002; Chen, 2007; van de Schoot, Lugtig & Hox, 2012). In all hierarchical levels for this study, change in goodness-of-fit indices available in SAS 9.4 (i.e., $\Delta \chi^2$, $\Delta CFI$ and $\Delta RMSEA$ and SRMR) was used to establish the extent of measurement invariance.

The question of measurement invariance was first addressed in comparing students at the upper and lower secondary levels. There were four secondary levels but each level comprised fewer than 200 responses. Hence, the sample was divided into two groups, upper and lower secondary level, for the MGCFA. Measurement invariance of this division would be likely to be the same as that for secondary one to four, given that a secondary one respondent can be of the same age as a secondary two respondent, and a secondary three respondent can be of the same age as a secondary four respondent. $\Delta CFI$ and $\Delta RMSEA$ values were computed based on the available fit indices from SAS 9.4. Results, presented in Table 25, suggest that there is measurement invariance between the upper and
lower secondary levels as the $\Delta CFI$ and $\Delta RMSEA$ values are within recommended thresholds.

Table 25. Test of measurement invariance by school level

<table>
<thead>
<tr>
<th>Measurement invariance model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$\chi^2$/df</th>
<th>$\Delta CFI$</th>
<th>$\Delta RMSEA$</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configural</td>
<td>856.33*</td>
<td>494</td>
<td>1.73</td>
<td>-</td>
<td>-</td>
<td>0.05</td>
</tr>
<tr>
<td>Metric</td>
<td>900.80*</td>
<td>546</td>
<td>1.65</td>
<td>0.00</td>
<td>0.01</td>
<td>0.06</td>
</tr>
<tr>
<td>Scalar</td>
<td>977.49*</td>
<td>572</td>
<td>1.71</td>
<td>0.01</td>
<td>0.01</td>
<td>0.06</td>
</tr>
<tr>
<td>Strict</td>
<td>1015.66*</td>
<td>624</td>
<td>1.63</td>
<td>0.00</td>
<td>0.01</td>
<td>0.06</td>
</tr>
</tbody>
</table>

*p < .001

Measurement invariance was then considered in terms of differences between males and females. Based on the same $\Delta CFI$ and $\Delta RMSEA$ thresholds, results, presented in Table 26, suggest that there is measurement invariance between male and female respondents. Together, these results suggest that there was measurement invariance found for the MRQ in this study and that results could be interpreted meaningfully.

Table 26. Test of measurement invariance by gender

<table>
<thead>
<tr>
<th>Measurement invariance model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$\chi^2$/df</th>
<th>$\Delta CFI$</th>
<th>$\Delta RMSEA$</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configural</td>
<td>836.55*</td>
<td>494</td>
<td>1.69</td>
<td>-</td>
<td>-</td>
<td>0.04</td>
</tr>
<tr>
<td>Metric</td>
<td>881.02*</td>
<td>546</td>
<td>1.61</td>
<td>0.00</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>Scalar</td>
<td>938.48*</td>
<td>572</td>
<td>1.64</td>
<td>0.01</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>Strict</td>
<td>999.16*</td>
<td>624</td>
<td>1.60</td>
<td>0.00</td>
<td>0.00</td>
<td>0.07</td>
</tr>
</tbody>
</table>

*p < .001

The number of participants by stream for the normal academic and technical stream (n = 138), and the number of participants from School X (n = 128) were less than the recommended minimum sample (200) required for
MGCFA. Hence, measurement invariance analyses by stream and school would not be meaningful.
Like many other countries in the developed world, Singapore recognises that her continued prosperity relies on developing future generations that are not only highly skilled from an academic perspective, but also, discerning in judgment, with a strong sense of right and wrong. The Singapore Ministry of Education (MOE), in recognition of this need, has now incorporated moral reasoning as a core element of its Character and Citizenship Education (CCE) syllabus. The CCE also identifies Kohlberg’s stage model explicitly as a key framework for conceptualising moral reasoning within this context.

As noted in Chapter 2, various instruments have appeared in the education research literature over the years to assess moral reasoning based on Kohlberg’s stage model. Most prominent amongst these are Kohlberg’s original Moral Judgement Interview (MJI), the original Defining Issues Test (DIT), and the Defining Issues Test 2 (DIT2). Despite the fact that each of the instruments that has appeared has been shown to produce useful information on the forms of moral reasoning used by individuals, none of these provide a practical means by which the development of moral reasoning can be monitored in large groups of secondary level students. As a result, teachers who wish to facilitate learning within this area have no means by which to monitor the success of their efforts, or target specific areas upon which to focus within their teaching practices.
The aim of the present study was to develop an instrument (the *Moral Reasoning Questionnaire*, or MRQ) to measure these attributes on a broad-scale basis within Singapore secondary classrooms. This final chapter summarises the results of the preliminary validation study conducted to evaluate the properties of the MRQ, and to make suggestions for the use of the instrument within schools. Directions for future research on the MRQ are also discussed.

5.1. SUMMARY OF RESULTS

The preliminary validation work on the MRQ presented in this thesis was designed to address three main aspects of Messick’s (1993, 1995) unified concept of validity. Messick suggested that, in validating a new instrument, five forms of evidence could be gathered to evaluate the validity of that instrument: (i) evidence related to the test content; (ii) evidence related to response processes; (iii) evidence related to the internal structure of the instrument; (iv) evidence related to associations between the instrument and other variables; and (v) evidence on the consequences of using the instrument. The study reported here provided evidence on the MRQ with respect to the first three of these criteria.

The forms of evidence collected were presented in Chapter 4. In general, this evidence provided both qualitative and quantitative support for the validity of the MRQ. The next section of this chapter provides a brief summary of the main findings with respect to the first three of Messick’s (1993, 1995) criteria.
With regard to evidence on the content validity of the MRQ (i.e., the items and the format), an extensive literature review was conducted to identify an appropriate and established moral reasoning framework (i.e., Kohlberg’s Theory of Moral Development) prior to the development of the instrument. Further support for the appropriateness of the MRQ in terms of test content was obtained through an expert panel group, which comprised education professionals from the Ministry of Education and schools in Singapore. The latter group reviewed each of the MRQ items in relation to attributes such as language and sentence structure, as well as the format of the MRQ, and appropriate revisions were incorporated within the instrument before any further validation work was done.

With respect to evidence on the response processes required by the MRQ, information on this aspect of the instrument was gathered by means of cognitive interviews with a small group of representative Singapore students. Based on the results of these interviews, there was no evidence to suggest that the interviewees were overwhelmed by cognitive load in completing the questionnaire, or that the item dilemmas or response options were ambiguous. There was also no evidence that any students used alternative methods (e.g., recalling certain frameworks or model answers) in addressing the questions posed by the MRQ. Thus, this preliminary investigation provided strong support for the validity of the MRQ from a response process perspective.

Finally, with respect to evidence on the internal structure of the MRQ, the results of the EFA, CFA and MGCFA analyses all yielded satisfactory goodness-of-fit indices, supporting the conclusion that the MRQ measures a single underlying
latent trait (moral reasoning). The results of the Rasch analyses further supported the unidimensionality of the MRQ, and also addressed issues of DIF within the MRQ items. Based on the triangulation of evidence across all of these analyses, the present study has provided strong support for the internal structure of the MRQ as an instrument.

5.2. IMPLICATIONS FOR PRACTICE

Based on the evidence presented in this thesis, the MRQ holds promise as an instrument that can be used in Singapore secondary schools to monitor students’ development in the area of moral reasoning. As an accessible instrument with sound psychometric properties, the MRQ would be suitable for use in a large-scale format (e.g., in entire secondary classrooms or schools). A further advantage of this instrument is that minimal training is required for teacher to administer and score the test. This adds further support to the notion that the MRQ can provide a practical means by which students’ development in moral reasoning can be monitored, hence addressing a major gap identified in this context.

The 2014 Standards for Educational and Psychological Testing manual states five non-mutually exclusive categories of uses that psychological tests can address: (i) testing for diagnosis; (ii) testing for neuropsychological evaluations; (iii) testing for intervention planning and outcome evaluation; (iv) testing for judicial and governmental decisions; and (v) testing for personal awareness, social identity, and psychological health, growth, and action. The MRQ could prove useful in addressing two of these goals (iii and v) in Singapore schools. It should be noted
here, however, that the intention of the MRQ is to measure moral reasoning development and not to evaluate students’ morality levels. As a developmental tool for personal growth and social awareness, students should be assisted to attain the desired moral reasoning developmental levels accorded by the MOE guidelines. The MRQ scores should not be used for other purposes such as placement or selection decisions (e.g., for student leadership appointments), or promote the labelling of students in terms of their ‘levels’ of moral reasoning.

Administered across different levels at a common time frame (e.g., end of the year), the MRQ may also provide schools with cumulative information on the moral reasoning development of entire cohorts of students within the school. Over the years, this information can be used, together with the knowledge of policy documents (e.g., age appropriate moral reasoning levels and teacher observations or reports), across a spectrum ranging from trend analysis for the planning of school-wide CCE curricular programmes, to individual analysis and feedback for the moral reasoning development of an individual student via conversations with the teacher. Koh (2009) suggested various intervention programmes schools can use to develop moral reasoning in Singapore secondary school students. Some of the more effective include: (i) cultural transmission, which involves the sharing of socially desirable cultural practices and traditions; (ii) story-telling, in which students identify and develop values through narratives; (iii) perspective-taking, which allows students to empathise with others and consider their needs; and (iv) community involvement, in which students volunteer to programs to carry out morally-grounded actions.
5.3. DIRECTIONS FOR FUTURE RESEARCH

By its nature, the “validation process never ends, as there is always additional information that can be gathered to more fully understand a test and the inferences that can be drawn from it” (AERA, APA & NCME, 2014, p. 21). Based on Messick’s five criteria for evaluating the validity of an instrument, further work is needed to validate the MRQ before it is used on a larger-scale basis.

First, with respect to the content of the MRQ, further studies should be undertaken to review the items and format under two suggested situations: (i) if the MRQ is used in a context other than Singapore secondary classrooms; and (ii) whether information and communication technology could permit the items to be presented differently (e.g., the moral dilemmas are presented as cartoon strips instead of a short paragraph) to broaden the use of the instrument to other age groups, demographic groups, and/or contexts.

Second, with respect to evaluating the response processes required by the MRQ, studies could be undertaken with a larger sample than the one used in the present study. In particular, if the MRQ is used beyond the Singapore context, cognitive interviews should be conducted with students in that context to ascertain the evidence supporting validity related to response processes. The 2014 Standards for Educational and Psychological Testing manual states that “tests selected for use in psychological testing should be suitable for the characteristics and background of the test taker” (AERA, APA, & NCME, 2014, p. 165). Pre-conceived response processes may differ for contexts beyond Singapore secondary schools.
Third, while this study provided favourable evidence on the validity of the the 26-item MRQ with respect to three of Messick’s (1993, 1995) evidentiary framework criteria, a large proportion (nearly 80%) of the participants were from the express stream and a similarly large proportion (nearly 75%) were female. Though the Rasch analyses suggested that inferences drawn from the measure would not be confounded by students’ demographic attributes (e.g., gender, school or stream), and that the MGCFA supported measurement invariance by gender, the disproportionate sample by gender and stream could optically suggest otherwise. Hence, a more representative sample could be invited to participate in subsequent studies. With more data, MGCFA could also be performed to establish measurement invariance by stream and school.

Further, the MRQ’s relations to other variables (the fourth form of evidence supporting validity under Messick’s unified concept of validity) were not evaluated in this study. Given the constraints under which the research work was done, and given the timeframe of the study, the participating schools did not wish to have students undertaking further measures alongside the MRQ. Future research, therefore, is needed to compare and correlate the MRQ with other variables. Two obvious choices would be the original MJI (Kohlberg, 1984) and the DIT2 (Rest et al., 1999). These analyses would provide direct evidence on how the MRQ scores align with, or depart from, those obtained through other instruments designed to assess moral reasoning based on Kohlberg’s model. In addition to these measures, a range of other variables (e.g., cognitive ability, personality, social attitudes) could also be used in the validation of the MRQ.
In addition to the above forms of validation, further evidence could be sought on the question of whether the MRQ scores differ in relation to sensitive variables such as cognition, culture and social economic status (SES). This also was not possible in the context of the present study, due to perceived sensitivity of these questions within the Singapore context. The Rasch analysis did indicate that there was no item bias for school, educational stream, gender and level within the sample. However, method bias due to cultural traits remains possible (Byrne, Oakland, Leong, van de Vijver, Hambleton, Cheung, & Bartram, 2009).

Some work is currently underway by the researchers to explore whether the MRQ can be used in other cultural contexts. In particular, an adapted version has already been developed for use in the context of Australian schools (the MRQ – Australia, or MRQ-A) (see Appendix C). The MRQ-A comprises 30 items that are similar to those in the original MRQ. To ensure that the construct (moral reasoning) to be measured by the items remain similar to the MRQ, everything including the instructions and item stimuli are the same. There are, however, slight changes to the wording of five items within the instrument. For example, the term ‘hawker centre’ was changed to ‘food hall’ and ‘curry puffs’ to ‘pies’ for items 28 and 30. A food hall in Australia is the equivalent of a hawker centre in Singapore; pies are also common snacks in Australia. Students would be more familiar to the terms ‘food hall’ and ‘pies’. Despite the slight changes, care was taken for the construct to be measured and item stimuli to be as similar to the original MRQ as possible. It is anticipated that, besides making the items more familiar to the Australian students to minimise construct-irrelevant sources of
variance, these slight changes would not alter the construct measured by the instrument. As the MRQ-A will serve to determine measurement invariance due to culture, all 30 items should be administered to the Australian students before further decisions are taken (e.g., to remove items that cause misfit).

With regard to exploring possible differences in responses to the MRQ based on cognitive ability, various previous studies have reported correlations in the .20 to .50 range between moral judgments and measures of intelligence, aptitude and achievement (Rest, 1979; Thoma & Dong, 2014). Thus, while correlations between measures of moral judgment (e.g., DIT P-scores) with measures of cognitive ability have not been high enough to conclude that moral reasoning is a proxy for cognitive ability, the intricacies and extent of relationships between these variables should be explored further with respect to the MRQ. As noted by Rest (1979):

> While related to cognition, moral judgment is moral; it is not merely the application of cognitive skills or intelligence to moral questions or situations. Measures of moral judgment correlate with measures of moral attitudes, choice, and behaviour to an extent not accounted for by IQ or other “pure” cognitive variables (p. 7).

Nevertheless, more could be done to establish the MRQ as a measure of moral reasoning independent of cognition and intelligence. This study had access to students from mainstream secondary schools and hence, the MRQ appears to be well fit for purpose in Singapore secondary school students. To ascertain that there is no DIF across a wider range of cognitive levels, and hence provide
assurance that students of higher aptitude do not exhibit higher levels of moral reasoning, the MRQ could be administered to students who are not in mainstream secondary schools, though have comparable literacy levels.

It should be noted here that the MRQ was developed specifically to minimise confounds by cognitive ability to the extent possible, through the use of simple English and short sentences that should be comprehensible to students within the targeted age range. The Rasch analysis also did not find DIF for educational stream, a proxy for cognition as students are streamed based on academic standards. Despite these preliminary findings, further empirical evidence on possible biases within the MRQ due to differences in cognitive ability need to be explored more thoroughly.

Messick’s (1993, 1995) fifth criterion for evaluating the validity of an instrument, the consequences of testing, was also not evaluated in this study, owing to the preliminary stage of the work, and time available to complete the validation. This criterion could be evaluated through a longitudinal study of the impact of the MRQ on practices after schools have adopted the instrument as a measure within the CCE. Backwash effects and unintended consequences can also be studied to establish how the MRQ may influence the teaching of moral values in the CCE curriculum. An unintended consequence could be the invidious labelling of students. One possibility for circumventing this could be to use standardised scores and band classifications (e.g., developing towards expectations, meeting expectations, exceeding expectations) rather than actual score values in reporting
results. At this stage, however, no data are available to suggest whether the use of the instrument will have positive or negative consequences within schools.

Beyond future research areas suggested by Messick’s unified concept of validity, the MRQ could potentially be developed further to take an adaptive test form, using computer adaptive testing (CAT). CAT is tailored testing that adapts to the ability of an individual who uses a computer to access the test. CAT has generated much interest and has been applied by the United States for their Graduate Record Examination and Computerized Placement Test, as well as being used by the National institute of Educational Measurement in The Netherlands (Latu & Chapman, 2002). Nonetheless, a search did not yield the use of CAT for assessments in the affective domain, which suggests an opportunity to provide such a tool via an MRQ (CAT) version.

Various reasons for switching from conventional paper-based testing to CAT were cited by van der Linden and Glas (2010): (i) the possibility for examinees to schedule tests at their convenience; (ii) tests are taken in a more comfortable setting and with fewer people around than in large-scale paper-and-pencil administrations; (iii) electronic processing of test data and reporting of scores are faster; and (iv) wider ranges of questions and test content can be put to use. In this regard and for the following reasons, the MRQ (CAT) would afford greater convenience to users of the instrument though more items should be developed: (i) the MRQ and its duration would generally be shorter than the current state; (ii) there would be better security as items might differ for an individual (with the assumption that the MRQ item bank is large and kept current); (iii) the MRQ
(CAT) can be implemented on-demand and faster, with immediate auto-scoring and hence reporting; (iv) a student who attempts the MRQ (CAT) would not be penalised for a “wrong” response because of the system would select appropriate items for him or her, and eventually measure his moral reasoning level more accurately; (v) the MRQ (CAT) might reduce test anxiety, if any, the system would allow any extra time as required by the student (Rudner, 1998); and (vi) students would not be confronted with items that are overly difficult or easy (Eggen & Straetmans, 2009).

CAT is normally done through a system. This system “adjusts the test by presenting easy questions to a low-proficiency examinee and difficult questions to a high-proficiency examinee” (Economides & Roupas, 2007). Essentially, the objective of CAT is to construct an optimal test for each examinee. To achieve this optimality, an examinee’s trait level, in this instance, moral reasoning development level, is estimated during the test and items appropriate to the examinee are selected from an item bank. Normally, an item is presented to the examinee and the first item for a test is of “medium difficulty for the total population” (Green, Bock, Humphreys, Linn & Reckase, 1984, p. 32). After the examinee responds, a score is noted and the next most appropriate item will be selected by the system from the item bank for the examinee. Figure 29 shows an example of how optimality is achieved by administering items that ultimately stabilise the ability estimate of the test taker. Items that are too difficult or easy will not be selected for an examinee based on his/her trait level, as these items do not provide much information about the examinee’s ability.
5.4. CONCLUSIONS

This study represents a first effort by the author to contribute to the field of measuring moral reasoning development in secondary level students. The impetus for the study was the lack of an instrument that could measure moral reasoning efficiently in a large scale setting, as well as the importance of this trait within the Singapore education system. Existing instruments, such as the MJI and DIT, have practical limitations as noted previously. Further, the moral dilemmas presented in these instruments will not necessarily be relevant to students at the secondary level in the Singapore context. To date, only one instrument by Soh (1987), the Test of Moral Values (TMV), has been developed for Singapore students. The TMV, however, is not supported by validation studies based on Messick’s (1993, 1995) validity criteria, and thus, its utility within this context remains questionable. While further research is needed to validate the instrument before it can be adopted on a widespread basis, the results reported in this thesis suggest that the MRQ holds considerable promise for use within this context.
References


Measuring Adolescents’ Moral Judgment: An Evaluation of the Socio-

moral Reflection Measure - Short Form Objective SRM-SFO.

Byrne, B.M., Oakland, T., Leong, F.T. L., van de Vijver, F.J.R., Hambleton, R.K.,
research and testing practices: Implications for improved education and
training in psychology. *Training and Education in Professional Psychology*,
3(2), 94-105.

464-504.

Cheung, C. (1999). Ethical judgment and ethical reasoning on business issues: A
cross-lag model for university students in Hong Kong. *College Student

testing measurement invariance, *Structural Equation Modeling: A
Multidisciplinary Journal*, 9(2), 233-255.

judgment reloaded: A moral dilemma validation study. *Frontiers in
Psychology*, 5, 1-18.


Development*, 48(1/2), 1-124.


in everyday life. Freeland, Oxfordshire, United Kingdom: Inter-Disciplinary Press.


http://dx.doi.org/10.4135/9781412983891.n13.


Thoma, S.J., Narvaez, D., Rest, J., Derryberry, P. (1999). Does moral judgment development reduce to political attitudes or verbal ability? Evidence


world: Key themes in inter-professional ethics (pp. 47-62). Lanham, Maryland: Lexington Books.
Appendix A1 – Approval by the MOE
EDUN N82-07-005

28 September 2015

Mr. Lyndon Lim

Dear Mr. Lim,

DEVELOPMENT AND VALIDATION OF A MORAL REASONING MEASURE FOR SINGAPOREAN SECONDARY SCHOOL STUDENTS

I refer to your application for approval to collect data from schools.

2. I am pleased to inform you that the Ministry has no objections to your request to conduct the research in 3 secondary schools, subjected to the following conditions:

   a) the approved research proposal is adhered to during the actual study in schools;
   b) the data collected is kept strictly confidential and used for the stated purpose only; and
   c) the findings are not published without written approval from the Ministry and a copy of the findings is shared with the Ministry upon completion of the study.

3. When conducting the data collection in the schools, please ensure that the following are carried out:

   a) consent is obtained from the Principals for the study to be conducted in the schools;
   b) written parental consent is obtained before conducting the study with the students;
   c) students and teachers are informed that participation in the study is voluntary and they do not need to provide any sensitive information (e.g. name and NRIC No.);
   d) participation by the schools are duly recorded in Annex A; and
   e) the data collection in schools is completed within 3 months from the date of this letter.

4. Please show this letter and all the documents included in this mail package (i.e. the application form, research proposal and research instrument(s) marked as seen by MOE) to seek approval from the Principals and during the actual study.

Yours sincerely,

Muhamad Imran bin Mohd Yusof
Research Analyst, Management Information/Corporate Research Office
Planning Division
for Permanent Secretary (Education)

Note to Principal: Please refer to MOE notification PA/25/12 for the Guidelines on Data Collection from Schools.
Appendix A2 – Letter to Principals
23 July 2015

Name & Address

Participant Information Form

Dear Principal's Name,

You are invited to participate in a research project titled "The Development and Validation of a Moral Reasoning Measure for Singapore Secondary School Students". This project will be conducted by Lyndon, a doctoral candidate with the Graduate School of Education.

As part of developing the measure, 3 to 5 teachers involved in planning character and citizenship education (CCE) programmes will be asked to review questionnaire items. They can do this at their own time within a 2-week period. Thereafter, 5 randomly selected students across secondary 1 to 4 will complete the questionnaire and be interviewed individually during this process. Each interview will take about 2 hours in your school.

The items will be refined and administered online to your students based on feedback from the teachers and 5 students. Students can expect to complete the questionnaire within 40 minutes. 80 students (2 classes) from each secondary level will be invited to complete the questionnaire.

There will be no risk involved to your teachers and students. Participation is voluntary and a participant is free to withdraw from the research at any time without prejudice in any way. The participant does not need to give any reason for withdrawing. Participation in this study does not prejudice any rights to compensation, which the student may have under statute or common law. Findings from this project would be published and presented but data will be anonymised.

This project has been approved by the Singapore Ministry of Education and the Human Research Ethics Office of the University of Western Australia. It is anticipated that this research will contribute to the professional practices of Singapore teachers engaged in CCE by providing a standardised measure of moral reasoning.

If you have any questions concerning the research project, please email Lyndon at lyndon.lim@research.uwa.edu.au or Dr. Elaine Chapman at elaine.chapman@uwa.edu.au.

I look forward to working with your school on this important research project.

Yours sincerely,

Lyndon Lim, Graduate Student

A/Prof. Elaine Chapman
Associate Dean, Research

Appendix to consent form: This research was approved by the University of Western Australia, in accordance with all ethical review and approval procedures. Any person considering participating in this research project, by agreeing to participate, may raise any question or issue with the researchers at any time.

In addition, any person or school with the agreement of respondents may make ethical issues of concern, and may raise any complaints about this research project to the Human Research Ethics Office at The University of Western Australia on 08 9380 3000 or email to humanresearch@uwa.edu.au.

All research participants are entitled to retain a copy of any Participant Information Sheet and/or Participant Consent Form relating to this research project.
Appendix A3 – Letter to Expert Panel
23 July 2015

Teacher-Participant Information Form

Dear Teacher,

I am a doctoral candidate at the Graduate School of Education, University of Western Australia working on a research project titled "The Development and Validation of a Moral Reasoning Measure for Singapore Secondary School Students".

As part of developing the measure, I will need you to review the items in the questionnaire based on (i) the comprehensibility of the item stem for 13 - 16 year old students (ii) the clarity of the options. You can do this anytime from XX Oct 2015 to XX Oct 2015. I will collect your reviews on XX Oct 2015.

Your review will be confidential and only my supervisor and I have access to it. It will be used to refine the items before they are administered to the students. There will be no risk involved to you. Participation is voluntary and you are free to withdraw from the research at any time without prejudice in any way.

This project has been approved by the Singapore Ministry of Education and the Human Research Ethics Office of the University of Western Australia. If you have any questions concerning the research project, please email me at lyndon.im@research.wwa.edu.au.

I look forward to working with you on this important research project.

Yours sincerely,

Lyndon Lim, Graduate Student

Approval to conduct this research has been provided by The University of Western Australia in accordance with its ethics review and approval procedures. Any person considering participation in this research project, or agreeing to participate, may raise any questions or concerns with the researchers at any time.

In addition, any person not satisfied with the response of researchers may raise ethics issues or concerns, and may make any complaints about this research project by contacting the Human Research Ethics Office at The University of Western Australia on (08) 6488 3763 or by emailing to hrecreview@uwa.edu.au.

All research participants are entitled to retain a copy of their Participant Information Sheet and/or Participant Consent Form, relating to this research project.
Teacher-Participant Consent Form

I (the participant) have read the information provided and any questions I have asked have been answered to my satisfaction. I agree to participate in this activity, realising that I may withdraw at any time without reason and without prejudice.

I understand that all identifiable (attributable) information that I provide is treated as strictly confidential and will not be released by the investigator in any form that may identify me. The only exception to this principle of confidentiality is if documents are required by law.

I have been advised as to what data is being collected, the purpose for collecting the data, and what will be done with the data upon completion of the research.

I agree that research data gathered for this study may be published provided my name or other identifying information is not used.

Participant name ___________________________ Date __________

Participant signature ________________________
Appendix A4 – Letters to Parents and Participants for Stage 2
23 July 2016

Parent Information Form

Dear Parent,

I am a doctoral candidate at the Graduate School of Education, University of Western Australia, working on a research project titled "The Development and Validation of a Moral Reasoning Measure for Singapore Secondary School Students".

Student responses to a moral reasoning questionnaire I am proposing are critical to the development and validation of this measure. Hence, I am seeking your consent for your child/ward to do this questionnaire and be interviewed by me. The process will be done during the end of year examinations and it will take about 2 hours.

Your child/ward’s involvement in this research is voluntary. S/he is free to withdraw at any stage without prejudice in any way. No reason is required for withdrawal. When you sign the consent form, I will assume that you agree to allow your child to participate and allow me to use the data in this research. Findings from this research will be published or presented but all data will be anonymised and your child will not be identifiable.

This project has been approved by the Singapore Ministry of Education and the Human Research Ethics Office of the University of Western Australia. If you have any questions concerning the research project, please email me at lyndon.lim@research.uwa.edu.au.

I look forward to your consent for your child/ward to work with you on this project.

Yours sincerely,

Lyndon Lim, Graduate Student
Parent Consent Form

I have read the information provided and any questions I have asked have been answered to my satisfaction. I agree to let my child/ward participate in this activity, realising that s/he may withdraw at any time without reason and without prejudice.

I understand that all identifiable (attributable) information that my child/ward provides is treated strictly confidential and will not be released by the investigator in any form that may identify him/her. The only exception to this principle of confidentiality is if documents are required by law.

I have been advised as to what data is being collected, the purpose for collecting the data, and what will be done with the data upon completion of the research.

I agree that research data gathered for the study may be published provided my child/ward's name or other identifying information is not used.

______________  ______________
Parent's signature  Date

Name of student

___________________________
School

___________________________
Class
23 July 2015

Participant Information Form

Dear Participant,

I am a doctoral candidate at the Graduate School of Education, University of Western Australia working on a research project titled “The Development and Validation of a Moral Reasoning Measure for Singapore Secondary School Students”.

As part of developing the measure, I will need you to complete the questionnaire. While you are doing the questionnaire, I will interview you on your thoughts. The whole process will take about 2 hours.

Your responses will be confidential and only accessible to me. They will be used to refine the items before they are administered to the other students. There will be no risk involved to you. Participation is voluntary and you are free to withdraw from the research at any time without prejudice in any way.

The project has been approved by the Research Ethics Office of the University of Western Australia. If you have any questions concerning the research project, please email me at lyndon.lim@research.uwa.edu.au.

I look forward to working with you on this important research project.

Yours sincerely,

Lyndon Lim, Graduate Student
Participant Consent Form

I (the participant) have read the information provided and any questions I have asked have been answered to my satisfaction. I agree to participate in this activity, realising that I may withdraw at any time without reason and without prejudice.

I understand that all identifiable (attributable) information I provide is treated as strictly confidential and will not be released by the investigator in any form that may identify me. The only exception to this principle of confidentiality is if documents are required by law.

I have been advised as to what data is being collected, the purpose for collecting the data, and what will be done with the data upon completion of the research.

I agree that research data gathered for the study may be published provided no names or other identifying information is not used.

Participant name ______________________ Date ______________________

Participant signature ______________________
Appendix A5 – Letters to Parents and Participants for Stage 3
23 July 2015

Parent Information Form

Dear Parent,

I am a doctoral candidate at the Graduate School of Education, University of Western Australia, working on a research project titled "The Development and Validation of a Moral Reasoning Measure for Singapore Secondary School Students".

Student responses to a moral reasoning questionnaire I am proposing are critical to the development and validation of this measure. Hence, I am seeking your consent for your child/ward to do the questionnaire. It will take your child/ward about 40 minutes to complete the questionnaire and it will be administered at the end of your examinations.

Your child/ward’s involvement in this research is voluntary. She is free to withdraw at any stage without prejudice in any way. No reason is required for withdrawal. When you sign the consent form, I will assume that you agree to allow your child to participate and allow me to use the data in this research. Findings from this research will be published or presented but all data will be anonymised and your child will not be identifiable.

This project has been approved by the Singapore Ministry of Education and the Human Research Ethics Office of the University of Western Australia. If you have any questions concerning the research project, please email me at lyndon.lim@research.uwa.edu.au.

I look forward to your consent for your child/ward to work with you on this project.

Yours sincerely,

Lyndon Lim, Graduate Student
Parent Consent Form

I have read the information provided and any questions I have asked have been answered to my satisfaction. I agree to let my child/ward participate in this activity, realising that s/he may withdraw at any time without reason and without prejudice.

I understand that all identifiable (attributable) information that my child/ward provides is treated as strictly confidential and will not be released by the investigator in any form that may identify him/her. The only exception to this principle of confidentiality is if documents are required by law.

I have been advised as to what data is being collected, the purpose for collecting the data, and what will be done with the data upon completion of the research.

I agree that research data gathered for the study may be published provided my child/ward's name or other identifying information is not used.

_________________________  ____________
Parent's signature  Date

Name of student

_________________________
School

_________________________
Class
23 July 2015

Participant Information Form

Dear Participant,

I am a doctoral candidate at the Graduate School of Education, University of Western Australia working on a research project titled "The Development and Validation of a Moral Reasoning Measure for Singapore Secondary School Students".

As part of developing the measure, I will need you to complete the web-based questionnaire. The whole process will take about 40 minutes in your school computer laboratory.

Your responses will be confidential and only my supervisor and I have access to it. It will be used to refine the items. There will be no risk involved to you. Participation is voluntary and you are free to withdraw from the research at any time without prejudice in any way.

This project has been approved by the Singapore Ministry of Education and the Human Research Ethics Office of the University of Western Australia. If you have any questions concerning the research project, please email me at lyndon.lim@research.uwa.edu.au.

I look forward to working with you on this important research project.

Yours sincerely,

Lyndon Lim, Graduate Student
Participant Consent Form

I (the participant) have read the information provided and any questions I have asked have been answered to my satisfaction. I agree to participate in this activity, realising that I may withdraw at any time without reason and without prejudice.

I understand that all identifiable (attributable) information that I provide is treated as strictly confidential and will not be released by the investigator in any form that may identify me. The only exception to this principle of confidentiality is if documents are required by law.

I have been advised as to what data is being collected, the purpose for collecting the data, and what will be done with the data upon completion of the research.

I agree that research data gathered for the study may be published provided my name or other identifying information is not used.

Participant name __________________________ Date __________________________

Participant signature ______________
Appendix A6 – Letter to Teachers and MRQ Administration Instructions
Teacher Information Form

Dear Participant,

I am a doctoral candidate at the Graduate School of Education, University of Western Australia working on a research project titled “The Development and Validation of a Moral Reasoning Measure for Singapore Secondary School Students”.

As part of developing the measure, students will need to complete a questionnaire. On the day of administering the questionnaire, I will need you to ensure that the students complete the questionnaire uninterruptedly in a computing laboratory. The whole process will take about 40 minutes. Before students start doing the questionnaire, please remind your students about the contents of the Participant Information Form, in particular:

1) The purpose of this research
2) That they are free to participate and withdraw from the research
3) That their information will be kept confidential.

If a student chooses not to be part of the research study or decides to withdraw at any time, please allow the student to do his homework in the computing laboratory without any repercussions whatsoever.

This project has been approved by the Singapore Ministry of Education and the Human Research Ethics Office of the University of Western Australia. If you have any questions concerning the research project, please email me at lyndon.lim@research.uwa.edu.au or my supervisor and Chief Investigator, Dr. Elaine Chapman at elaine.chapman@uwa.edu.au.

I look forward to working with you on this important research project.

Yours sincerely,

Lyndon Lim, Graduate Student

Approval to conduct this research has been provided by The University of Western Australia, in accordance with its ethics review and approval procedures. Any person considering participation in this research project, or agreeing to participate, may raise any questions or concerns with the researchers at any time.

In addition, any person not satisfied with the response of researchers may raise ethics issues or concerns, and may make any complaints about this research project by contacting the Human Research Ethics Office at The University of Western Australia on (08) 6488 3163 or by emailing lyndon.lim@uwa.edu.au

All research participants are entitled to retain a copy of any Participant Information Sheet and/or Participant Consent Form relating to this research project.
Dear Teacher,

Please read the following instructions to your students before they do the questionnaire. Please leave this sheet of paper on this table after reading it to the students for the next teacher to use.

If students have any questions, please call me immediately at [phone number]. I will come to the lab.

Students,

You are about to do an online questionnaire comprising 30 questions. In doing so, you will help to refine the questionnaire that will hopefully be used to assess moral reasoning.

Each question in the questionnaire has 2 stages. In stage 1, you will read a short scenario and select one out of two options. You will then be shown stage 2 depending on what you chose in stage 1. In stage 2, you will need to drag and drop three options depending on the order of importance in helping you decide your option in stage 1. The option that you drag and drop to the top will be the most important and the one you place at the bottom will be the least important. You can click on “next” or “back” buttons to change your options anytime but only the final options will be recorded by the system.

You have 45 minutes to complete the questionnaire. As your name is not required, you will not be identified. Just be honest when answering the questions. Please do not rush through it. If you feel unwell, you can stop doing the questionnaire. Once you have completed the questionnaire, please read your storybook. Please do the questionnaire only once.

You may begin now. The password is “integrity”.
Appendix B – Moral Reasoning Questionnaire
Instructions

Dear Student,

Please read the following notes before clicking "NEXT". By clicking "NEXT", you agree and consent to participate in this questionnaire.

**Notes**

1) Please complete all the 30 items in this questionnaire.

2) Each item has 2 parts, **A** and **B**. In part **A**, you need to choose 1 of 2 options after reading a scenario. Depending on your choice, you will be directed to either part **B1** or **B2** where you will need to rank 3 options in order of importance to you in deciding to choose the option in part **A**.

3) You can click on "BACK" or "NEXT" to edit your options at any time to view the options in both stages.

4) There is no time limit but please complete the questionnaire within this single session.

**NAME**

D1. Gender

D2. Age

D3. Class

D4. Please select your school from the list below.

Q1

1a. You are running for president of the Students' Council, of which you are currently treasurer. The current president is struggling with her school work but is also running for presidency. At the elections, you see one of your friends whom you know hasn’t been attending council meetings regularly and is not supposed to vote. After the elections, this friend tells you he voted for you. The next day, the results are announced and you win by one vote. You recall that the friend who was not supposed to vote voted for you. You would...

- report the truth.
- not report the truth.

1b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report the truth. You would...

- report the truth as you might be disqualified for the presidency if you were found for hiding the truth.
- report the truth as it is your duty to do so as the current treasurer, and it is against the Students' Council rule to hide the truth.
- report the truth as it is important that the elections are conducted fairly.

1b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report the truth. You would...

- not report the truth as you would benefit more being a president than a treasurer. You believe you can be a better president than the current one.
- not report the truth as you think that the current president will need more time to focus on her school work next year.
- not report the truth as everyone has the right to pursue his/her goal.

Q2

2a. You are a prefect. During a test, you saw your best friend cheating. This is not the first time it has happened. You know that the parents of this friend have high expectations and your friend will be caned for poor results or if they knew about the cheating. Your best friend would definitely hate you for reporting the cheating. You would...

- report the cheating.
- not report the cheating.
2b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report the cheating. You would...

- report the cheating as you might be removed from the prefectural board if you were found for not reporting it.
- report the cheating as you want your friend to learn what is right.
- report the cheating as it is important to be fair to everyone who took the test.

2b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report the cheating. You would...

- not report the cheating as you would not want your best friend to hate you.
- not report the cheating as you don’t want your best friend to be caned.
- not report the cheating as everyone has a right to do what is best for himself.

Q3

3a. Your friend and you are in the same class. Your friend is very weak in Biology but better than you in other subjects. You just found out that she copied a Biology essay online and submitted it for an assignment. She asked you to keep quiet about it. This assignment is important and if she fails it, she will be retained for a year and she would be sad. You would...

- report that your friend copied the essay.
- not report that your friend copied the essay.

3b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report the copying. You would...

- report the matter to the teacher as she might give you a zero for the essay if she found out that you ignored the act of copying.
- report the matter as copying is not allowed according to the school rules.
- report the matter as it is important that everyone in the class is assessed fairly.

3b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report the copying. You would...
Q4

4a. You are in the school badminton team. Your friend is as good as you in badminton and has won many matches for the school. Both of you are competing for the position of the badminton team captain. In the promotional exam, you saw your friend cheat. After asking him, you found that if he failed the exam, his parents would stop him from playing badminton. He asks you to keep it to yourself. You would...

- report the cheating.
- not report the cheating.

4b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report the cheating. You would...

- report the cheating to the teacher. Hiding the truth from the teacher might cost you the position of team captain.
- report the cheating to the teacher as school rules state that cheating is a serious offence.
- report the cheating as it is more important that everyone is treated fairly for the exam.

4b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report the cheating. You would...

- not report the cheating to the teacher as your teammates would be unhappy with you if they knew you did.
- not report the cheating to the teacher as your friend asked you to keep it to yourself.
- not report the cheating as your friend has a right to pursue his interest in badminton even if it meant cheating for the promotional exam.

Q5
5a. You found out that 2 of your good friends insulted the discipline master on their blogs. The discipline master is known to be a fair and good teacher. When you asked them, they admitted but asked you to keep it to yourself. 2 days later, your principal called you into his office and asked whether you knew the culprits behind the offensive blogs. You would...

- reveal that you knew the culprits.
- not reveal that you knew the culprits.

5b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to reveal that you knew the culprits. You would...

- reveal to the principal that you knew but kept quiet so that you won't get into further trouble.
- reveal your friends' names as the school rules state that bad behaviour should not be ignored.
- reveal your friends' names as it is more important to address such bad behaviour than your relationship with your friends and classmates.

5b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to reveal that you knew the culprits. You would...

- not reveal your friends' names as it would be bad if other classmates knew that you gave your friends' name to the principal.
- not reveal your friends' names as your friends asked you to keep it a secret so that they wouldn't get into trouble.
- not reveal your friends' names. The fact that your friends admitted but asked you to keep it a secret indicates that they know it was unethical to publish the insults. It is more important that they realise their mistake than to let the principal know and punish them.

Q6

6a. In the class online chat, you noticed some of your friends talking bad about a classmate and they seem to be enjoying it. You would...

- report this matter.
- not report this matter.

6b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report this matter. You would...
207

- report it to the teacher so that you might not face detention class if the online chat was found out.
- report this issue to the teacher as the school encourages us not to badmouth others.
- report this if your friends continue talking bad about the classmate. It is more important that your friends learn to respect others.

6b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report this matter. You would...

- not report the remarks as they may not be true and more importantly, it wouldn’t benefit you if your classmates found out you reported it.
- not report this as everyone seems to be enjoying the chat. After all, the chat topic will soon die off.
- not report this as everyone has the right to discuss about their opinions of others.

Q7

7a. During recess, you realised you forgot to do your Mathematics assignment. With only 15 minutes left to the next lesson, your best friend offers you his assignment to copy as he doesn’t want you to be scolded. Your Mathematics teacher is strict and will definitely inform your parents if you admitted you forgot to do the assignment as this is not your first time. You would...

- copy the assignment.
- not copy the assignment.

7b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to copy the assignment. You would...

- copy your best friend’s assignment and submit it so that you won’t be scolded.
- copy your friend’s assignment as the teacher does not allow any late submission and school rules state the assignments should be submitted on time.
- copy your best friend’s assignment and submit it on time so that your parents and teacher would not be disappointed with you.

7b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to copy the assignment. You would...
• not copy your best friend’s assignment as you might be punished if you were found out.
• not copy your best friend’s assignment as it is against the school rule to copy assignments.
• not copy but do what you can within 15 minutes and submit the assignment as it is more important to be honest.

Q8

8a. You are in a debate team and your class is expecting your team to win. Your friend is on an opposing team and the debate will take place tomorrow. You sit beside this friend and can see his debate notes. His notes could help your team win. You would...
- copy his notes.
- not copy his notes.

8b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to copy his notes. You would...
- copy his notes but use it for yourself instead of sharing it entirely with your team so that you might have a chance to be the best speaker.
- copy and share his notes with your team to have a better chance of winning.
- copy his notes and share with your team as both teams have been given equal opportunities to prepare for the debate whatever it takes. It is a fair fight.

8b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to copy his notes. You would...
- not copy his notes as you might be punished and disqualified if you were found out.
- not copy his notes as the school does not allow copying without acknowledging the source.
- not copy his notes nor take a look as both teams should have equal chances to win the debate.

Q9

9a. You are a prefect standing at the school gate to record the names of latecomers. The gate is just behind the assembly area. You best friend John arrives at the gate just as you shut the gate.
You know he could be late as his parents’ quarrel affected him as he told you over the phone last night. You would...

- allow John in.
- not allow John in.

9b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to allow John in. You would...

- allow John in by pretending to look away and not shutting the gate so that you won’t be reprimanded by the Prefect Master for letting a latecomer in.
- allow John in. If your classmates see you allowing John in, they wouldn’t dislike you.
- allow John in as he wasn’t late intentionally. It wasn’t really his fault.

9b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to allow John in. You would...

- not allow John in as you might be punished by the teacher if she saw it.
- not allow John in as you are duty bound as a prefect.
- not allow John in as it is important to be fair to every latecomer.

Q10

10a. The class beside yours took the same Science test 2 hours ago. Over recess, you heard your friend Jane asking what the questions were. Even though this is just a class test, Jane says that she needs whatever help to boost her failing results. Jane normally sits beside you during a test. You would...

- report Jane to the teacher.
- not report Jane to the teacher.

10b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report Jane to the teacher. You would...

- report Jane to the teacher as you might be given a zero if your teacher found out that you knew but didn’t report the matter.
- report Jane to the teacher as the school does not allow cheating.
• report Jane to the teacher as it is more important to ensure fairness in the test for everyone.

10b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report Jane to the teacher. You would...

• not report but ask Jane what the questions were and discuss the answers with her so that both she and you would benefit and she would value you as a friend more.
• not report but ask Jane to share the questions with the class so that everyone in the class can do well.
• not report Jane as this could be a morale booster for her to work harder in future tests.

Q11

11a. You are the group leader of a class project. A high percentage of the grade is based on group work. Your good friend is in the same group but she is not contributing even though you have asked her to, likely due to some family issues. You would...

- report your friend to the teacher.
- not report your friend to the teacher.

11b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report your friend to the teacher. You would...

• report your friend to the teacher for fear of being accused and later made to apologise to the class for being biased.
• report your friend to the teacher as the teacher warned everyone about contributing actively to the group.
• report your friend to the teacher as everyone else in the group has a right to get a better grade than her. At the same time, explain to the teacher that your friend had some family issues.

11b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report your friend to the teacher. You would...

• not report your friend to the teacher so that she will not dislike you.
• not report your friend to the teacher as your group didn’t seem unhappy about it. If you did, your group might say you are not a good friend.
Q12

12a. You are the class monitor. Your teacher tasked you to divide the class equally and randomly into groups for project work. You know there are some classmates who are notorious and will not contribute to a group. You would...
- divide the class randomly.
- divide the class randomly with conditions.

12b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to divide the class randomly. You would...
- divide the class randomly for fear of being accused and later punished by the teacher for being biased.
- divide the class randomly as the teacher instructed you to do so.
- divide the class randomly as it is important to be fair to everyone.

12b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to divide the class randomly with conditions. You would...
- divide the class randomly but for your own group you would ensure that members would only be those you can work with.
- divide the class randomly based on your classmates’ preferences so that they would be happy with you.
- divide the class randomly but ensure that each group has a classmate known to be notorious. Everyone should learn how to work with difficult people.

Q13

13a. You are the umpire (referee) for an inter-school badminton match between School A and B. At match point, the line judge (who determines whether the shuttlecock lands in or outside the court) says that School A has won. Spectators all shout that School B has won. You can decide whether School A or B wins. If School A wins, their next match would be against your school and your school would likely lose as School A is stronger. You would...
- decide School B as the winner.
13b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding School B as the winner. You would...

- decide School B as the winner as you don’t want the spectators to criticise you for making a wrong decision.
- decide School B as the winner. This way, your teammates would thank and not blame you.
- decide School B as the winner. After all, you are the umpire and your decision should be undisputed.

13b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding School A as the winner. You would...

- make it clear to everyone that you are respecting the line judge’s decision that School A has won. This way, you won’t be blamed for anything.
- respect the line judge’s decision as the game should be conducted this way.
- decide School A as the winner but ask School B whether they want another point to be played in the name of sportsmanship and fairness.

Q14

14a. Your teacher asked you to connect her laptop to the projector and then left the class for a while. While setting up the laptop, you saw your final exam marks on it. You noticed that your mark was 1 short of an A1 grade. An A1 grade could make a difference on your streaming for the following year and your parents will scold you if you don’t make it to the top stream. You would...

- amend the marks on the laptop.
- not amend the marks on the laptop.

14b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to amend the marks on the laptop. You would...

- change the mark on the laptop so that you won’t be scolded by your parents.
- not touch the laptop in case your classmates disapproved of your action.
- add 1 mark to everyone in the class so that everyone gets 1 mark more. This would be fairer to everyone.
14b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to change the marks on the laptop. You would...

- not change your mark in case your teacher finds out and you are punished.
- not touch the laptop as it is against the school rule that states that one should be honest and upright.
- close the laptop screen and ensure that the laptop was not meddled with until the teacher returned as it is important to be fair to everyone.

Q15

15a. You are a prefect. You see your close friend who always helps you in your homework smoking in the toilet but he asks you to keep it to yourself. This is not his first time. You know that this classmate has abusive parents and if you report him, he would be suspended and his parents will hit him. You would...

- report him to the teacher.
- not report him to the teacher.

15b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report him to the teacher. You would...

- report him. If you don’t report him and the teacher finds out, you might be reprimanded and lose the position of prefect.
- report him to the teacher as you are duty bound as a prefect and the school rules do not allow smoking in school.
- report him as it is more important that he leads a healthier lifestyle.

15b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report him to the teacher. You would...

- not report him so that you won’t lose him as a close friend.
- not report him so that his parents would not abuse him or feel hurt. He would also not feel worried about suspension.
- not report him but ask him to admit to the teacher that he smoked in the toilet and ask the teacher for help to kick the habit.
Q16

16a. Your class has been asked to give feedback about your Science teacher using a feedback form. You are required to write your name on the feedback form for identification. The school requires that you be truthful in this feedback. Many classmates and you feel that this teacher is biased and impatient but most of your classmates intend to give nice and false feedback. You would...

- give true feedback.
- not give true feedback.

16b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to give true feedback. You would...

- give true feedback so that you won't be criticised by anyone for being dishonest.
- give true feedback as the school values honesty.
- give true feedback as it is more important for the teacher to know the truth and improve.

16b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to give true feedback. You would...

- give false and nice feedback so that you would not be marked by the teacher.
- give false and nice feedback so that you will not feel different from your peers.
- give false feedback as this teacher might only be hurt unnecessarily if everyone gives harsh and true feedback.

Q17

17a. You are the Chairman of a uniform group. The Vice-Chairman just left the uniform group as he transferred to another school. Your teacher is asking you for a name to fill the position of Vice-Chairman. Your close friend Samy is qualified and he told you he wants to assume the position. Yet, there is another schoolmate who is more qualified than Samy. Your teacher will likely accept whichever name you suggest as she is new to the school. You would...

- suggest Samy's name.
- not suggest Samy's name.
17b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to suggest Samy's name. You would...

- suggest Samy's name as the Vice-Chairman. If Samy becomes Vice-Chairman, you won't have any trouble working with him as you know him well.
- suggest Samy's name as the Vice-Chairman as Samy is your good friend and he told you he wants to be Vice-Chairman.
- suggest Samy's name as the Vice-Chairman. You know that Samy can co-lead the uniform group well with you and bring it to greater heights.

17b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to suggest Samy's name. You would...

- not suggest Samy's name to avoid any scolding by the teacher if she finds out that you mentioned Samy's name because he is your close friend.
- not suggest Samy's name as you know there is a more qualified schoolmate and the most qualified person should be recommended.
- not suggest Samy's name as it is more important to be fair to everyone and the uniform group.

Q18

18a. You just received your report book and noticed you got an A1 for Mathematics, different from the A2 grade you got for the actual paper. Apparently, your teacher keyed A1 into the report book by mistake. Some friends sitting around you saw the A1 grade and said it was definitely an error; they said you should just let it be in case their good grades were also errors made by the teacher. An A1 grade for Mathematics will help you get into the Science stream and your parents will be delighted. You would...

- report the error.
- not report the error.

18b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report the error. You would...

- report the error to the teacher. If she found out later that you knew but didn’t report the error, you would be scolded for being dishonest.
- report the error to the teacher as the school rules state that one must be honest.
• report the error to the teacher as it is important that the grading is accurate and fair to everyone.

18b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report the error. You would...

• not report the error as it wasn’t your fault and you wouldn’t be blamed for the error.

• not report the error since your friends told you to let it be. You wouldn’t want to be blamed for any change in grade for your friends’ grades if the teacher found other mistakes in the class’s results.

• not report the error. You feel that there’s more benefit to leave the Mathematics grade as A1; you can get into the Science stream and your parents will be happy.

Q19

19a. You work part-time in a fairly expensive restaurant with your classmates. Your job is to wash dishes. After working in the restaurant for just a day, you are thoroughly disgusted. You have seen that the kitchen is very dirty and has an infestation of cockroaches. You mention this to the restaurant manager who just smiles and walks away. You would...

- report the matter to the authorities.
- not report the matter to the authorities.

19b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report the matter. You would...

• report the matter to the relevant authorities quickly so that you won’t be criminalised if the dirty kitchen was found out by authorities.

• report the matter to the relevant authorities as the law does not permit such unhygienic practices.

• report the matter as it is important that all diners consume clean food.

19b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report the matter. You would...

• not report the matter to the relevant authorities so that you won’t be scolded by the restaurant manager.
Q20

20a. Your best friend works part-time in a mobile phone shop. When you visit him, he asks whether you want a mobile phone that you always wanted. He says that the store somehow has one extra mobile phone and insists on giving one to you so that he can tally the stocks. You would...

- accept the phone.
- not accept the phone.

Q20b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to accept the phone. You would...

- accept the mobile phone as that was what you always wanted. After all, it is an extra mobile phone.
- accept the mobile phone as your friend wants to give it to you and he is your friend.
- help him tally the stocks first. If it is really an extra, you would take it.

Q20b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to accept the phone. You would...

- not accept the mobile phone as you do not want to get into any unnecessary trouble.
- not accept the mobile phone as it is against the law to tally stocks by giving extra phones away.
- ask him to keep the mobile phone for himself or someone else who needs it more than you.

Q21

21a. You are unprepared for a test today. To buy time, you consider feigning illness so that you can take it another day when you have more confidence. You know the teacher will not let you take the test another day unless you have a valid reason. You would...

- feign illness to skip the test.
21b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to feign illness and skip the test. You would...
- feign illness to skip the test, have more time to prepare for it and score better than the rest.
- feign illness to skip the test. This way, you can have more preparation time and your parents or teacher won’t be disappointed with you for not doing well for the exam.
- feign illness to skip the test and have more time to prepare for it as it doesn’t hurt anyone to do so.

21b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to feign illness and skip the test. You would...
- not feign illness to skip the test as you might be punished if you were found out.
- not feign illness as it is against the school rule to lie and skip the test without a valid reason.
- not feign illness to skip the test as it is important to be fair to everyone.

Q22

22a. Some of your friends consume alcohol on their 16th birthday parties. You attend many of these parties. Your parents do not allow you to drink but you tell them it is difficult to reject your friends at so many parties. Though your parents ask you to tell your friends you can’t drink due to medical reasons, you feel the real reason is that they don’t want to be seen as over-protective. You would...
- take alcoholic drinks.
- not take alcoholic drinks.

22b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to take alcoholic drinks. You would...
- drink and enjoy the company but keep it from your parents. If your parents found out, they might ground you for a few days.
- drink and enjoy the company at selected birthday parties and explain your intention to your parents. This way, both your friends and parents won’t get totally upset.
- drink as you don’t see anything wrong with drinking as long as you do not overdo it.
22b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to take alcoholic drinks. You would...

- not drink no matter how to avoid scolding and punishment by your parents.
- tell your friends you won’t drink as the legal age for alcohol consumption is 18.
- tell your friends you can’t drink as your parents are concerned with your health.

Q23

23a. You failed your test. Your parents are angry and order you to stay home. This evening, they are away for a wedding dinner and you know they won’t be back till late night. Your friends ask you out for dinner and really hope you will join them. You know that your parents will likely not find out if you went out and you feel that being grounded because you failed a test is unfair. You would...

- go for dinner with your friends.
- not go for dinner with your friends.

23b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to go for dinner with your friends. You would...

- go for dinner with your friends and come home way before your parents are home so that you won’t get into trouble.
- go for dinner with your friends so that your friends won’t be unhappy.
- go for dinner with your friends because it is not right to ground you because you failed a test.

23b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to go for dinner with your friends. You would...

- not go for the dinner in case your parents find out and punish you.
- not go for dinner as you have to abide by house rules set by your parents.
- not go for dinner because you can stay home and study and not disappoint your parents.

Q24
24a. Your brother spent all his money on gaming. Just as he asks you for money to buy textbooks as his exams are next week, you find a pouch containing $500. There is a name inside the pouch but you do not know the owner. You would...

- report the lost pouch to the police.
- not report the lost pouch to the police.

24b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report the lost pouch to the police. You would...

- report the lost pouch to the police in case someone saw you taking it and you could get into trouble.
- report the lost pouch to the police as taking without permission is similar to stealing which is against the law.
- report the lost pouch to the police as the owner may need the money urgently and source for alternative funding for your brother.

24b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report the lost pouch to the police. You would...

- keep the $500 for your brother as he would be grateful to you.
- keep the $500 for your brother as he needs it urgently for his exams.
- take $250 for your brother and report the remaining amount and lost pouch to the police. At least, you would be happy that you were able to help your brother.

Q25

25a. You just bought a new handphone (e.g. iPhone 6+) using your parents’ money. As you were walking out of the shop, you accidentally dropped it and the screen cracked. No one saw you drop the phone. You would...

- return the handphone and tell the shop you found a crack on it. Ask for a one-for-one exchange.
- not return the handphone. Just live with the crack.

25b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to return the handphone. You would...
return the smartphone, say you found a crack on it and ask for a new one so that your parents won't scold you for being clumsy.

return the smartphone, say you found a crack on it and ask for a new one so that you wouldn't be wasting your parents' hard-earned money.

return the smartphone, say you found a crack on it and ask for a new one as everyone has the right to make his/her hard-earned money worth and well-spent.

25b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to return the smartphone. You would...

- not return the smartphone so that your parents won't scold you for being dishonest.
- not return the smartphone as insisting that you found a crack on it and asking for a new one is unlawful and can be considered as cheating.
- not return the smartphone as it is more important to be fair to the store that sold you the smartphone.

Q26

26a. Your parents do not earn much. You just bought a textbook from the school bookstore and noticed that the cashier returned you more change than she was supposed to. You would...

- return the excess change.
- not return the excess change.

26b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to return the excess change. You would...

- return the excess change as you do not want to be caught and sent for detention class for being dishonest.
- return the excess change as it is against the school rules to be dishonest.
- return the excess change as it is more important to be honest than to save some money.

26b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to return the excess change. You would...

- keep the excess change since nobody would find out you took it and you won't get into any trouble.
Q27

27a. Your friends and you join a science quiz. Depending on how many questions are answered correctly, the winning team will be awarded up to $10000 and they can give it a charitable organisation of their choice. 2 days before the quiz, you find a list of questions for the quiz on the table at the venue. Your team has been struggling to learn the science concepts included in the quiz. You would...

- note the list of questions.
- ignore the list of questions.

27b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to note the list of questions. You would...

- note the list of questions but not tell your team about it so that it doesn’t reflect poorly on you. During the quiz, you would try to answer as many questions as possible so that your team can win most of the $10000.
- tell your team the questions so that the team won’t feel so stressed about preparing for the quiz.
- note the list of questions and get your team to win most of the $10000 for the sake of charity. It is more important to win more than less for charity.

27b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to ignore the list of questions. You would...

- ignore the list of questions as you might be caught and disqualifed for cheating even if it is for charity.
- ignore the list of questions as this is against competition rules.
- ignore the list of questions as it is important to be fair to the other team and the prize sponsor.
28a. You are at a hawker centre. You see an old man selling curry puffs by the stairway. He isn’t licensed to sell but still does it for a living. His curry puffs could be dirty and cause food poisoning. You would...

- report the old man to the authorities.
- not report the old man to the authorities.

28b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report the old man to the authorities. You would...

- report the old man to the authorities as you wouldn’t want to be criticised for not reporting such unlicensed practices to authorities.
- report the old man to the authorities as it is not right for people to sell food without a license by law.
- report the old man as you would be helping other potential customers avoid eating the dirty curry puffs.

28b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report the old man to the authorities. You would...

- not report the old man as you normally don’t eat his curry puffs and you don’t want to be in any unnecessary trouble.
- not report the old man so that you won’t be seen by others as a heartless person.
- not report the old man but speak to him and offer help in other ways like referring him for financial aid.

Q29

29a. You are 5 minutes to being late for school. Just when you are about to enter the school gate, you see an old man fall down. There is no one at the school gate to help him except you. The nearest person is about 100 metres away at the bus stop. You have been late for school before and you know that if you are late for school this time, you will be punished. You would...

- help the old man.
- not help the old man.

29b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to help the old man. You would...
• help the old man as you do not want the public at the bus stop to criticise you for being unhelpful.

• help the old man as you are wearing the school uniform and what you do reflects on the school.

• help the old man even if you were late because his safety is more important.

29b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not help the old man. You would...

• not help the old man as you don’t want to be punished for being late. Besides, others will see and come and help him.

• not help the old man as the school rules require students to be punctual.

• not help the old man as you do not want to disappoint my parents and teachers in being late again.

Q30

30a. You often eat with your family at a hawker centre. Each time you eat there, a few people will approach you and sell packets of tissue paper even though you have tissue packets. Some of these people are handicapped or old and others look young. You know that these people are not licensed to sell tissue paper. You would...

○ buy tissue paper from anyone who approaches you.

○ not buy tissue paper from anyone who approaches you.

30b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to buy tissue paper from anyone who approaches you. You would...

• buy tissue from anyone who approaches as you do not want people around to criticise you for being unhelpful.

• buy tissue from anyone who approaches you as they will be happy that someone is buying from them.

• buy tissue from anyone who approaches you as they are just earning a living.

30b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to buy tissue paper from anyone who approaches you. You would...
• not buy tissue from anyone who approaches you as you might get into trouble if authorities see you supporting unlicensed sellers.

• not buy tissue from anyone who approaches you as the law doesn’t allow unlicensed sellers.

• not buy tissue from anyone who approaches you so that licensed sellers such as shops around the hawker centre have a chance to earn a living.
Appendix C – MRQ (Australia)
Instructions

Dear Student,

Please read the following notes before clicking "NEXT". By clicking "NEXT", you agree and consent to participate in this questionnaire.

Notes

1) Please complete all the 30 items in this questionnaire.

2) Each item has 2 parts, A and B. In part A, you need to choose 1 of 2 options after reading a scenario. Depending on your choice, you will be directed to either part B1 or B2 where you will need to rank 3 options in order of importance to you in deciding to choose the option in part A.

3) You can click on "BACK" or "NEXT" to edit your options at any time to view the options in both stages.

4) There is no time limit but please complete the questionnaire within this single session.

NAME

D1. Gender

D2. Age

D3. Class
D4. Please select your school from the list below.

Q1

1a. You are running for president of the Students’ Council, of which you are currently treasurer. The current president is struggling with her school work but is also running for presidency. At the elections, you see one of your friends whom you know hasn’t been attending council meetings regularly and is not supposed to vote. After the elections, this friend tells you he voted for you. The next day, the results are announced and you win by one vote. You recall that the friend who was not supposed to vote voted for you. You would...
   report the truth.
   not report the truth.

1b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report the truth. You would...
   report the truth as you might be disqualified for the presidency if you were found for hiding the truth.
   report the truth as it is your duty to do so as the current treasurer, and it is against the Students’ Council rule to hide the truth.
   report the truth as it is important that the elections are conducted fairly.

1b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report the truth. You would...
   not report the truth as you would benefit more being a president than a treasurer. You believe you can be a better president than the current one.
   not report the truth as you think that the current president will need more time to focus on her school work next year.
   not report the truth as everyone has the right to pursue his/her goal.
Q2

2a. You are a school councillor. During a test, you saw your best friend cheating. This is not the first time it has happened. You know that the parents of this friend have high expectations and your friend will be caned for poor results or if they knew about the cheating. Your best friend would definitely hate you for reporting the cheating. You would...

report the cheating.
not report the cheating.

2b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report the cheating. You would...

report the cheating as you might be removed from the Students' Council if you were found for not reporting it.
report the cheating as you want your friend to learn what is right.
report the cheating as it is important to be fair to everyone who took the test.

2b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report the cheating. You would...

not report the cheating as you would not want your best friend to hate you.
not report the cheating as you don't want your best friend to be caned.
not report the cheating as everyone has a right to do what is best for himself.

Q3

3a. Your friend and you are in the same class. Your friend is very weak in Biology but better than you in other subjects. You just found out that she copied a Biology essay online and submitted it for an assignment. She asked you to keep quiet about it. This assignment is important and if she fails it, she will be retained for a year and she would be sad. You would...

report that your friend copied the essay.
not report that your friend copied the essay.

3b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report the copying. You would...

report the matter to the teacher as she might give you a zero for the essay if she found out that you ignored the act of copying.
report the matter as copying is not allowed according to the school rules.
report the matter as it is important that everyone in the class is assessed fairly.

3b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report the copying. You would...

not report the matter so that your friend won’t hate you.
not report the matter as you should be sincere in the friendship and your friend asked you to keep it a secret.
not report the matter as it is everyone’s right to strive to be promoted.

Q4

4a. You are in the school badminton team. Your friend is as good as you in badminton and has won many matches for the school. Both of you are competing for the position of the badminton team captain. In the promotional exam, you saw your friend cheat. After asking him, you found that if he failed the exam, his parents would stop him from playing badminton. He asks you to keep it to yourself. You would...
report the cheating.
not report the cheating.

4b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report the cheating. You would...
report the cheating to the teacher. Hiding the truth from the teacher might cost you the position of team captain.

report the cheating to the teacher as school rules state that cheating is a serious offence.

report the cheating as it is more important that everyone is treated fairly for the exam.

4b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report the cheating. You would...

not report the cheating to the teacher as your teammates would be unhappy with you if they knew you did.

not report the cheating to the teacher as your friend asked you to keep it to yourself.

not report the cheating as your friend has a right to pursue his interest in badminton even if it meant cheating for the promotional exam.

Q5

5a. You found out that 2 of your good friends insulted the discipline master on their blogs. The discipline master is known to be a fair and good teacher. When you asked them, they admitted but asked you to keep it to yourself. 2 days later, your principal called you into his office and asked whether you knew the culprits behind the offensive blogs. You would...

reveal that you knew the culprits.

not reveal that you knew the culprits.

5b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to reveal that you knew the culprits. You would...

reveal to the principal that you knew but kept quiet so that you won’t get into further trouble.

reveal your friends’ names as the school rules state that bad behaviour should not be ignored.

reveal your friends’ names as it is more important to address such bad behaviour than your relationship with your friends and classmates.
5b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to reveal that you knew the culprits. You would...

- not reveal your friends’ names as it would be bad if other classmates knew that you gave your friends’ name to the principal.
- not reveal your friends’ names as your friends asked you to keep it a secret so that they wouldn’t get into trouble.
- not reveal your friends’ names. The fact that your friends admitted but asked you to keep it a secret indicates that they know it was unethical to publish the insults. It is more important that they realise their mistake than to let the principal know and punish them.

6a. In the class online chat, you noticed some of your friends talking bad about a classmate and they seem to be enjoying it. You would...
- report this matter.
- not report this matter.

6b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report this matter. You would...
- report it to the teacher so that you might not face detention class if the online chat was found out.
- report this issue to the teacher as the school encourages us not to badmouth others.
- report this if your friends continue talking bad about the classmate. It is more important that your friends learn to respect others.

6b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report this matter. You would...
- not report the remarks as they may not be true and more importantly, it wouldn’t benefit you if your classmates found out you reported it.
not report this as everyone seems to be enjoying the chat. After all, the chat topic will soon die
off.

not report this as everyone has the right to discuss about their opinions of others.

Q7

7a. During recess, you realised you forgot to do your Mathematics assignment. With only 15 minutes left
to the next lesson, your best friend offers you his assignment to copy as he doesn't want you to be
scolded. Your Mathematics teacher is strict and will definitely inform your parents if you admitted you
forgot to do the assignment as this is not your first time. You would...

copy the assignment.

not copy the assignment.

7b1. Drag and drop the following options in order of importance (1=most important and 3=least
important) when deciding to copy the assignment. You would...

copy your best friend's assignment and submit it so that you won't be scolded.

copy your friend's assignment as the teacher does not allow any late submission and school
rules state the assignments should be submitted on time.

copy your best friend's assignment and submit it on time so that your parents and teacher
would not be disappointed with you.

7b2. Drag and drop the following options in order of importance (1=most important and 3=least
important) when deciding not to copy the assignment. You would...

not copy your best friend's assignment as you might be punished if you were found out.

not copy your best friend's assignment as it is against the school rule to copy assignments.

not copy but do what you can within 15 minutes and submit the assignment as it is more
important to be honest.

Q8

8a. You are in a debate team and your class is expecting your team to win. Your friend is on an opposing team and the debate will take place tomorrow. You sit beside this friend and can see his debate notes. His notes could help your team win. You would...

copy his notes.
not copy his notes.

8b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to copy his notes. You would...

copy his notes but use it for yourself instead of sharing it entirely with your team so that you might have a chance to be the best speaker.
copy and share his notes with your team to have a better chance of winning.
copy his notes and share with your team as both teams have been given equal opportunities to prepare for the debate whatever it takes. It is a fair fight.

8b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to copy his notes. You would...

not copy his notes as you might be punished and disqualified if you were found out.
not copy his notes as the school does not allow copying without acknowledging the source.
not copy his notes nor take a look as both teams should have equal chances to win the debate.

Q9

9a. You are a school councillor standing at the school gate to record the names of latecomers. The gate is just behind the assembly area. You best friend John arrives at the gate just as you shut the gate. You know he could be late as his parents’ quarrel affected him as he told you over the phone last night. You would...

allow John in.
not allow John in.
9b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to allow John in. You would...

allow John in by pretending to look away and not shutting the gate so that you won’t be reprimanded by the Discipline Master for letting a latecomer in.
allow John in. If your classmates see you allowing John in, they wouldn’t dislike you.
allow John in as he wasn’t late intentionally. It wasn’t really his fault.

9b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to allow John in. You would...

not allow John in as you might be punished by the teacher if she saw it.
not allow John in as you are duty bound as a prefect.
not allow John in as it is important to be fair to every latecomer.

Q10

10a. The class beside yours took the same Science test 2 hours ago. Over recess, you heard your friend Jane asking what the questions were. Even though this is just a class test, Jane says that she needs whatever help to boost her failing results. Jane normally sits beside you during a test. You would...
report Jane to the teacher.
not report Jane to the teacher.

10b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report Jane to the teacher. You would...

report Jane to the teacher as you might be given a zero if your teacher found out that you knew but didn’t report the matter.
report Jane to the teacher as the school does not allow cheating.
report Jane to the teacher as it is more important to ensure fairness in the test for everyone.
10b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report Jane to the teacher. You would...

- not report but ask Jane what the questions were and discuss the answers with her so that both she and you would benefit and she would value you as a friend more.
- not report but ask Jane to share the questions with the class so that everyone in the class can do well.
- not report Jane as this could be a morale booster for her to work harder in future tests.

Q11

11a. You are the group leader of a class project. A high percentage of the grade is based on group work. Your good friend is in the same group but she is not contributing even though you have asked her to, likely due to some family issues. You would...

- report your friend to the teacher.
- not report your friend to the teacher.

11b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report your friend to the teacher. You would...

- report your friend to the teacher for fear of being accused and later made to apologise to the class for being biased.
- report your friend to the teacher as the teacher warned everyone about contributing actively to the group.
- report your friend to the teacher as everyone else in the group has a right to get a better grade than her. At the same time, explain to the teacher that your friend had some family issues.

11b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report your friend to the teacher. You would...

- not report your friend to the teacher so that she will not dislike you.
not report your friend to the teacher as your group didn’t seem unhappy about it. If you did, your group might say you are not a good friend.

not report your friend to the teacher as she could be affected by family issues and hence wasn’t contributing actively.

Q12

12a. You are the class monitor. Your teacher tasked you to divide the class equally and randomly into groups for project work. You know there are some classmates who are notorious and will not contribute to a group. You would…

divide the class randomly.

divide the class randomly with conditions.

12b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to divide the class randomly. You would…

divide the class randomly for fear of being accused and later punished by the teacher for being biased.

divide the class randomly as the teacher instructed you to do so.

divide the class randomly as it is important to be fair to everyone.

12b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to divide the class randomly with conditions. You would…

divide the class randomly but for your own group you would ensure that members would only be those you can work with.

divide the class randomly based on your classmates’ preferences so that they would be happy with you.

Everyone should learn how to work with difficult people.
Q13

13a. You are the umpire (referee) for an inter-school badminton match between School A and B. At match point, the line judge (who determines whether the shuttlecock lands in or outside the court) says that School A has won. Spectators all shout that School B has won. You can decide whether School A or B wins. If School A wins, their next match would be against your school and your school would likely lose as School A is stronger. You would...

decide School B as the winner.

decide School A as the winner.

13b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding School B as the winner. You would...

decide School B as the winner as you don't want the spectators to criticise you for making a wrong decision.

decide School B as the winner. This way, your teammates would thank and not blame you.

decide School B as the winner. After all, you are the umpire and your decision should be undisputed.

13b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding School A as the winner. You would...

make it clear to everyone that you are respecting the line judge's decision that School A has won. This way, you won't be blamed for anything.

respect the line judge's decision as the game should be conducted this way.

decide School A as the winner but ask School B whether they want another point to be played in the name of sportsmanship and fairness.

Q14

14a. Your teacher asked you to connect her laptop to the projector and then left the class for a while. While setting up the laptop, you saw your final exam marks on it. You noticed that your mark was 1 short of an A1 grade. An A1 grade could make a difference on your streaming for the following year and your parents will scold you if you don't make it to the top stream. You would...

amend the marks on the laptop.
not amend the marks on the laptop.

14b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to amend the marks on the laptop. You would...

- change the mark on the laptop so that you won’t be scolded by your parents.
- not touch the laptop in case your classmates disapproved of your action.
- add 1 mark to everyone in the class so that everyone gets 1 mark more. This would be fairer to everyone.

14b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to change the marks on the laptop. You would...

- not change your mark in case your teacher finds out and you are punished.
- not touch the laptop as it is against the school rule that states that one should be honest and upright.
- close the laptop screen and ensure that the laptop was not meddled with until the teacher returned as it is important to be fair to everyone.

Q15

15a. You are a student councillor. You see your close friend who always helps you in your homework smoking in the toilet but he asks you to keep it to yourself. This is not his first time. You know that this classmate has abusive parents and if you report him, he would be suspended and his parents will hit him. You would...

- report him to the teacher.
- not report him to the teacher.

15b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report him to the teacher. You would...
report him. If you don't report him and the teacher finds out, you might be reprimanded and lose the position of student councillor.

report him to the teacher as you are duty bound as a prefect and the school rules do not allow smoking in school.

report him as it is more important that he leads a healthier lifestyle.

15b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report him to the teacher. You would...

not report him so that you won't lose him as a close friend.

not report him so that his parents would not abuse him or feel hurt. He would also not feel worried about suspension.

not report him but ask him to admit to the teacher that he smoked in the toilet and ask the teacher for help to kick the habit.

Q16

16a. Your class has been asked to give feedback about your Science teacher using a feedback form. You are required to write your name on the feedback form for identification. The school requires that you be truthful in this feedback. Many classmates and you feel that this teacher is biased and impatient but most of your classmates intend to give nice and false feedback. You would...

give true feedback.
	not to give true feedback.

16b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to give true feedback. You would...

give true feedback so that you won't be criticised by anyone for being dishonest.

give true feedback as the school values honesty.

give true feedback as it is more important for the teacher to know the truth and improve.
16b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to give true feedback. You would...

- give false and nice feedback so that you would not be marked by the teacher.
- give false and nice feedback so that you will not feel different from your peers.
- give false feedback as this teacher might only be hurt unnecessarily if everyone gives harsh and true feedback.

Q17

17a. You are the Chairman of a uniform group. The Vice-Chairman just left the uniform group as he transferred to another school. Your teacher is asking you for a name to fill the position of Vice-Chairman. Your close friend Samy is qualified and he told you he wants to assume the position. Yet, there is another schoolmate who is more qualified than Samy. Your teacher will likely accept whichever name you suggest as she is new to the school. You would...

- suggest Samy’s name.
- not suggest Samy’s name.

17b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to suggest Samy’s name. You would...

- suggest Samy’s name as the Vice-Chairman. If Samy becomes Vice-Chairman, you won’t have any trouble working with him as you know him well.
- suggest Samy’s name as the Vice-Chairman as Samy is your good friend and he told you he wants to be Vice-Chairman.
- suggest Samy’s name as the Vice-Chairman. You know that Samy can co-lead the uniform group well with you and bring it to greater heights.

17b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to suggest Samy’s name. You would...

- not suggest Samy’s name to avoid any scolding by the teacher if she finds out that you mentioned Samy’s name because he is your close friend.
not suggest Samy’s name as you know there is a more qualified schoolmate and the most qualified person should be recommended.

not suggest Samy’s name as it is more important to be fair to everyone and the uniform group.

Q18

18a. You just received your report book and noticed you got an A1 for Mathematics, different from the A2 grade you got for the actual paper. Apparently, your teacher keyed A1 into the report book by mistake. Some friends sitting around you saw the A1 grade and said it was definitely an error; they said you should just let it be in case their good grades were also errors made by the teacher. An A1 grade for Mathematics will help you get into the Science stream and your parents will be delighted. You would...

report the error.
not report the error.

18b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report the error. You would...

report the error to the teacher. If she found out later that you knew but didn’t report the error, you would be scolded for being dishonest.

report the error to the teacher as the school rules state that one must be honest.

report the error to the teacher as it is important that the grading is accurate and fair to everyone.

18b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report the error. You would...

not report the error as it wasn’t your fault and you wouldn’t be blamed for the error.

not report the error since your friends told you to let it be. You wouldn’t want to be blamed for any change in grade for your friends’ grades if the teacher found other mistakes in the class’s results.

not report the error. You feel that there’s more benefit to leave the Mathematics grade as A1; you can get into the Science stream and your parents will be happy.
Q19

19a. You work part-time in a fairly expensive restaurant with your classmates. Your job is to wash dishes. After working in the restaurant for just a day, you are thoroughly disgusted. You have seen that the kitchen is very dirty and has an infestation of cockroaches. You mention this to the restaurant manager who just smiles and walks away. You would...

report the matter to the authorities.

not report the matter to the authorities.

19b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report the matter. You would...

report the matter to the relevant authorities quickly so that you won’t be criminalised if the dirty kitchen was found out by authorities.

report the matter to the relevant authorities as the law does not permit such unhygienic practices.

report the matter as it is important that all diners consume clean food.

19b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report the matter. You would...

not report the matter to the relevant authorities so that you won’t be scolded by the restaurant manager.

not report the matter otherwise your classmates may dislike you for causing them a loss of income should the restaurant be forced to close.

not report it but do your best to clean the kitchen as it is more important that the kitchen be properly cleaned than authorities clamping down on the restaurant business.

Q20

20a. Your best friend works part-time in a mobile phone shop. When you visit him, he asks whether you want a mobile phone that you always wanted. He says that the store somehow has one extra mobile phone and insists on giving one to you so that he can tally the stocks. You would...

accept the phone.
not accept the phone.

20b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to accept the phone. You would...

accept the mobile phone as that was what you always wanted. After all, it is an extra mobile phone.
accept the mobile phone as your friend wants to give it to you and he is your friend.
help him tally the stocks first. If it is really an extra, you would take it.

20b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to accept the phone. You would...

not accept the mobile phone as you do not want to get into any unnecessary trouble.
not accept the mobile phone as it is against the law to tally stocks by giving extra phones away.
ask him to keep the mobile phone for himself or someone else who needs it more than you.

Q21

21a. You are unprepared for a test today. To buy time, you consider feigning illness so that you can take it another day when you have more confidence. You know the teacher will not let you take the test another day unless you have a valid reason. You would...

feign illness to skip the test.
not feign illness to skip the test.

21b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to feign illness and skip the test. You would...

feign illness to skip the test, have more time to prepare for it and score better than the rest.
feign illness to skip the test. This way, you can have more preparation time and your parents or teacher won’t be disappointed with you for not doing well for the exam.

feign illness to skip the test and have more time to prepare for it as it doesn’t hurt anyone to do so.

21b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to feign illness and skip the test. You would...

not feign illness to skip the test as you might be punished if you were found out.
not feign illness as it is against the school rule to lie and skip the test without a valid reason.
not feign illness to skip the test as it is important to be fair to everyone.

Q22

22a. Some of your friends consume alcohol on their 16th birthday parties. You attend many of these parties. Your parents do not allow you to drink but you tell them it is difficult to reject your friends at so many parties. Though your parents ask you to tell your friends you can’t drink due to medical reasons, you feel the real reason is that they don’t want to be seen as over-protective. You would...
take alcoholic drinks.
not take alcoholic drinks.

22b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to take alcoholic drinks. You would...

drink and enjoy the company but keep it from your parents. If your parents found out, they might ground you for a few days.
drink and enjoy the company at selected birthday parties and explain your intention to your parents. This way, both your friends and parents won’t get totally upset.
drink as you don’t see anything wrong with drinking as long as you do not overdo it.
22b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to take alcoholic drinks. You would...

- not drink no matter how to avoid scolding and punishment by your parents.
- tell your friends you won't drink as the legal age for alcohol consumption is 18.
- tell your friends you can't drink as your parents are concerned with your health.

Q23

23a. You failed your test. Your parents are angry and order you to stay home. This evening, they are away for a wedding dinner and you know they won't be back till late night. Your friends ask you out for dinner and really hope you will join them. You know that your parents will likely not find out if you went out and you feel that being grounded because you failed a test is unfair. You would...

- go for dinner with your friends.
- not go for dinner with your friends.

23b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to go for dinner with your friends. You would...

- go for dinner with your friends and come home way before your parents are home so that you won't get into trouble.
- go for dinner with your friends so that your friends won't be unhappy.
- go for dinner with your friends because it is not right to ground you because you failed a test.

23b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to go for dinner with your friends. You would...

- not go for the dinner in case your parents find out and punish you.
- not go for dinner as you have to abide by house rules set by your parents.
- not go for dinner because you can stay home and study and not disappoint your parents.
Q24

24a. Your brother spent all his money on gaming. Just as he asks you for money to buy textbooks as his exams are next week, you find a pouch containing $500. There is a name inside the pouch but you do not know the owner. You would...

report the lost pouch to the police.

not report the lost pouch to the police.

24b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report the lost pouch to the police. You would...

report the lost pouch to the police in case someone saw you taking it and you could get into trouble.

report the lost pouch to the police as taking without permission is similar to stealing which is against the law.

report the lost pouch to the police as the owner may need the money urgently and source for alternative funding for your brother.

24b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report the lost pouch to the police. You would...

keep the $500 for your brother as he would be grateful to you.

keep the $500 for your brother as he needs it urgently for his exams.

take $250 for your brother and report the remaining amount and lost pouch to the police. At least, you would be happy that you were able to help your brother.

Q25

25a. You just bought a new handphone (e.g. iPhone 6+) using your parents' money. As you were walking out of the shop, you accidentally dropped it and the screen cracked. No one saw you drop the
phone. You would...

return the handphone and tell the shop you found a crack on it. Ask for a one-for-one exchange.
not return the handphone. Just live with the crack.

25b1. Drag and drop the following options in order of importance (1=most important and 3=least
important) when deciding to return the handphone. You would...

return the handphone, say you found a crack on it and ask for a new one so that your parents
won’t scold you for being clumsy.

return the handphone, say you found a crack on it and ask for a new one so that you wouldn’t
be wasting your parents' hard-earned money.

return the handphone, say you found a crack on it and ask for a new one as everyone has the
right to make his/her hard-earned money worth and well-spent.

25b2. Drag and drop the following options in order of importance (1=most important and 3=least
important) when deciding not to return the handphone. You would...

not return the handphone so that your parents won’t scold you for being dishonest.
not return the handphone as insisting that that you found a crack on it and asking for a new
one is unlawful and can be considered as cheating.
not return the handphone as it is more important to be fair to the store that sold you the
handphone.

Q26

26a. Your parents do not earn much. You just bought a textbook from the school bookstore and noticed
that the cashier returned you more change than she was supposed to. You would...

return the excess change.
not return the excess change.

26b1. Drag and drop the following options in order of importance (1=most important and 3=least
important) when deciding to return the excess change. You would...

return the excess change as you do not want to be caught and sent for detention class for being dishonest.

return the excess change as it is against the school rules to be dishonest.

return the excess change as it is more important to be honest than to save some money.

26b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to return the excess change. You would...

keep the excess change since nobody would find out you took it and you won't get into any trouble.

use the excess change to buy another textbook as you can help your parents save money and they would be happy.

use the excess change to make a donation to the needy.

Q27

27a. Your friends and you join a science quiz. Depending on how many questions are answered correctly, the winning team will be awarded up to $10000 and they can give it to a charitable organisation of their choice. 2 days before the quiz, you find a list of questions for the quiz on the table at the venue. Your team has been struggling to learn the science concepts included in the quiz. You would...

note the list of questions.

ignore the list of questions.

27b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to note the list of questions. You would...

note the list of questions but not tell your team about it so that it doesn't reflect poorly on you. During the quiz, you would try to answer as many questions as possible so that your team can win most of the $10000.

tell your team the questions so that the team won't feel so stressed about preparing for the quiz.
note the list of questions and get your team to win most of the $10000 for the sake of charity. It is more important to win more than less for charity.

27b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to ignore the list of questions. You would...

ignore the list of questions as you might be caught and disqualified for cheating even if it is for charity.
ignore the list of questions as this is against competition rules.
ignore the list of questions as it is important to be fair to the other team and the prize sponsor.

Q28

28a. You are at a food hall. You see an old man selling pies by the stairway. He isn’t licensed to sell but still does it for a living. His pies could be dirty and cause food poisoning. You would...

report the old man to the authorities.
not report the old man to the authorities.

28b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to report the old man to the authorities. You would...

report the old man to the authorities as you wouldn’t want to be criticised for not reporting such unlicensed practices to authorities.
report the old man to the authorities as it is not right for people to sell food without a license by law.
report the old man as you would be helping other potential customers avoid eating the dirty pies.

28b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to report the old man to the authorities. You would...
not report the old man as you normally don’t eat his pies and you don’t want to be in any unnecessary trouble.

not report the old man so that you won’t be seen by others as a heartless person.

not report the old man but speak to him and offer help in other ways like referring him for financial aid.

Q29

29a. You are 5 minutes to being late for school. Just when you are about to enter the school gate, you see an old man fall down. There is no one at the school gate to help him except you. The nearest person is about 100 metres away at the bus stop. You have been late for school before and you know that if you are late for school this time, you will be punished. You would...

help the old man.

not help the old man.

29b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to help the old man. You would...

help the old man as you do not want the public at the bus stop to criticise you for being unhelpful.

help the old man as you are wearing the school uniform and what you do reflects on the school.

help the old man even if you were late because his safety is more important.

29b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not help the old man. You would...

not help the old man as you don’t want to be punished for being late. Besides, others will see and come and help him.

not help the old man as the school rules require students to be punctual.

not help the old man as you do not want to disappoint my parents and teachers in being late again.
Q30

30a. You often eat with your family at a food hall. Each time you eat there, a few people will approach you and sell packets of tissue paper even though you have tissue packets. Some of these people are handicapped or old and others look young. You know that these people are not licensed to sell tissue paper. You would...

buy tissue paper from anyone who approaches you.
not buy tissue paper from anyone who approaches you.

30b1. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding to buy tissue paper from anyone who approaches you. You would...

buy tissue from anyone who approaches as you do not want people around to criticise you for being unhelpful.
buy tissue from anyone who approaches you as they will be happy that someone is buying from them.
buy tissue from anyone who approaches you as they are just earning a living.

30b2. Drag and drop the following options in order of importance (1=most important and 3=least important) when deciding not to buy tissue paper from anyone who approaches you. You would...

not buy tissue from anyone who approaches you as you might get into trouble if authorities see you supporting unlicensed sellers.
not buy tissue from anyone who approaches you as the law doesn't allow unlicensed sellers.
not buy tissue from anyone who approaches you so that licensed sellers such as shops around the food hall have a chance to earn a living.