Assessing moral judgment maturity using the Defining Issues Test (DIT-2) and the Sociomoral Reflection Measure- Short Form Objective (SRM-SFO)

Carien Boessenkool (2834766)

Master's thesis Clinical Child and Adolescent Psychology

Utrecht University

June 2022

Supervisors: Jan Boom & Daniel Brugman

Second grader: Iris Tjaarda

Abstract

The Defining Issues Test (DIT-2) and the Sociomoral Reflection Measure – Short Form Objective (SRM-SFO) aim to assess moral reasoning maturity. Both instruments are derived from the Neo-Kohlbergian theories of, respectively, James Rest and John Gibbs. This study, for the first time, compared both theories and their measures, using a sample of Dutch adolescents between 16- and 21-year-old ($N = 79$). Convergent validity of both moral maturity scores was assessed through correlations with theoretically relevant variables (age, gender, educational level, and self-reported anti-social behavior). In the SRM-SFO, convergent validity was supported by relationships between moral reasoning and age, gender, and anti-social behavior. The DIT-2 presented poor convergent validity and related only weakly to gender and educational level. Furthermore, this study compared moral maturity scores of both measures to each other, revealing a positive relationship, consistent with our expectation. Despite some limitations, this research enhances the understanding of the relationship between the theories of Rest and Gibbs and its measures. We conclude that the -easier to use- SRM-SFO is the better alternative in measuring moral judgment maturity in this age group and supports its use in large-scale research from young adolescence onward.

Keywords: moral judgment, moral development, objective measure, DIT-2, SRM-SFO

Assessing moral judgment maturity using the Defining Issues Test (DIT-2) and the Sociomoral Reflection Measure- Short Form Objective (SRM-SFO)

Morality: a study of philosophy or science? The answer to this question has often led to polarization between both parties (Maxim, 2014). The first psychologist to focus his research on the concept of morality and thereby break the philosophical tradition, was Lawrence Kohlberg (Blasi, 1990). Kohlberg used the cognitive developmental approach to examine moral reasoning. He applied Piaget's (1960) structural approach to the development of cognitions about the physical world, to the development of moral reasoning. According to Kohlberg, moral judgment development can be seen as a progression through a standard sequence of stages, where a person uses progressively more adequate forms of moral reasoning (Kohlberg, 1984). These days, some critics regard his work as outdated and beyond repair, and others formulated modifications. In this study, two Neo-Kohlbergian theories and their instruments were compared to each other, based on the following literature and quantitative analyses.

From Piaget to Kohlberg

In Piaget's theory (1932, 1965), cognitive skills are actively built, where the child starts with the development of memory and language in the first stage and progressively works its way up to logical and abstract thinking in the final stage. In general, this means that as children get older, they move through phases from primitive to more evolved thinking (Gibbs et al., 1992). According to Piaget, the development of moral judgment shows a progression from external morality based on consequences (punishment) and physical appearance at phase 1, to a more pragmatic and internal morality, with consideration of psychological contexts, at phase 2 (Basinger et al., 1995). This led to Piaget's formulation of two types of morality: a morality of constraint (unilateral) and a morality of cooperation (bilateral). Piaget argued that the two types of moral thought are not to be formed in two clear-cut stages, but that they tend to overlap. In the child, both types of morality exist, and in the adult, it is "simply a question of the proportions in which they are mixed" (Piaget, 1965, p. 85). Piaget used the context of children playing games to investigate this approach, particularly boys playing the game of marbles. His justification for this context is that rules of social games are elaborated, and passed on by children only, in contrast to moral rules as learned from adults, (Piaget 1932, 1965).

Kohlberg followed Piaget in rejecting moral development as a 'simple transmission' of moral rules from parents to children, because it does not explain how moral norms arise in the first place (Carpendale, 2003). Though, Kohlberg based his view on stages on Piaget's theory of cognitive development, rather than Piaget's phases of moral judgment development. In Kohlberg's theory, moral judgment develops through a sequence of six stages. These are identified in three levels of moral development: pre-conventional, conventional, and post-conventional, where each level includes two stages of reasoning (Kohlberg, 1984).

At the primary preconventional level, an individual understands 'right' and 'wrong' in terms of the consequences of action (e.g., reward or punishment) or authoritarian imposition. In stage 1, the good- or badness of an action is determined by its physical consequences. Key values are avoidance of punishment and unconditional respect for power. In stage 2, right action is defined by the satisfaction of the individuals' own needs. This includes elements of equal sharing and fairness, interpreted in terms of pragmatic or physical consequences ('an eye for an eye, a tooth for a tooth').

As the individuals' moral reasoning progresses to the second level, the conventional level, the maintenance of expectations by the family, group or nation becomes more important. Stage 3 emphasizes behavior that helps or pleases others, where the individual tries to gain approval from others. There is great emphasis for conformity to acceptable behavior and stereotypes. When the individual reaches stage 4, (s)he takes the perspective of a generalized member of the society. Within this perspective, adherence of societal, legal, and religious procedures, applied to all member of the society, is emphasized.

When the postconventional level of moral maturity is reached, the individual has the urge to define moral values apart from authoritarian power. Rather than rigidly following laws, stage 5 emphasizes rational considerations that can lead to the possibility of changing laws by social contract. In stage 6, the individual defines right by its conscience and self-chosen ethical principles that appeal to universality, comprehensiveness, and consistency. With this hypothesis about a fifth and sixth stage, Kohlberg went beyond the moral judgment as studied by Piaget. Piaget's theory of moral judgment is limited to the development of the schoolchild (from about 5 to about 11 years of age), while Kohlberg's theory concerns lifespan development (Colby & Kohlberg, 1987).

According to Kohlberg, individuals progress through this series of stages in an invariant sequence. The development can be accelerated, slowed down or stopped by cultural factors, but this does not change the sequence (Colby & Kohlberg, 1987). A stage is formed by a coherent pattern, that provides a general description of the individual's moral thought. Thus, a given response on a task represents an underlying thought organization. In Kohlberg's theory, lower stages are integrated and displaced by higher stages in a broader perspective. Once an individual would have evolved to a certain stage of moral reasoning, this way of thinking would be applied to all moral conflicts encountered. This means that stages must fulfill the criterion of consistency, which is implied by the notion of a 'structured whole' (Kohlberg, 1986). Kohlberg did acknowledge that there are exceptions, where people do not use their highest stage of moral reasoning, but he expected the general moral thinking across different contexts to be consistent with the stage (Carpendale, 2003; Kohlberg & Kramer, 1969).

The Moral Judgment Interview

To test this theory of moral development, Kohlberg developed the Moral Judgment Interview (MJI). The MJI involves interviewing a subject, using a series of moral conflicts. In the classical

dilemma of Heinz, the conflict is between the value of preserving life (by stealing a high-priced drug for his dying wife), or the value of upholding the law (and letting the wife die) (Carpendale, 2000). After the dilemmas have been presented orally by the interviewer, the interviewee answers several open-ended, probing questions. These are designed to elicit information regarding the moral reasoning that is used, to resolve the dilemma. Distinctly different moral rationales are expressed by subjects and captured in different stages of the moral development. The questions are explicitly prescriptive, to create normative judgments abouts what one *should* do, rather than predictive judgments about what one *would* do. To code a subject's response into a stage score, a detailed specification of concrete stage criteria and the moral concepts had to be used. The coding was a time-consuming, sometimes problematic, 17-steps process. This process includes the breaking down of the interview text into judgments, the matching of these judgments with previous standardized arguments in the scoring manual, and assigning the right stage score (Colby et al., 1983; Colby & Kohlberg, 1987).

However, in a twenty-year longitudinal study of Kohlberg and colleagues (Colby et al., 1983), using MJI, the highest stages were far less prevalent than other stages (Gibbs, 2003). The results did not necessarily constitute a 'universal theory' of moral judgment development, which led to revision and refinement in the formulation and scoring of stages. Kohlberg has always had difficulties analyzing the structure of the sixth stage, which is why he chose to eliminate this stage from his scoring manual. The justification for this decision is that the sixth stage would be too theoretical, to be mapped through interviews. Later, however, he came back to this (Kohlberg et al., 1990). Kohlberg's theory has been criticized and used as a source of inspiration by several people over the years, some of which will be discussed in the following part of this thesis.

A neo-Kohlbergian approach: Rest

Following Kohlberg, James Rest proposed an adapted model of moral judgment development. Rest's theoretical foundation starts with the idea of 'social justice'. Individuals are born into groups of people and must balance their own interests in social cooperation and achieve equilibrium in that balance. According to this foundation, moral judgement is developed through rights and responsibilities in a social system (Rest, 1978; Rawls, 1971). Thus, Rest laid greater emphasis on the role of others in moral development, next to the individual (Elm & Weber, 1994). The core ideas are Kohlberg's, but with some liberties. For instance, Rest uses the term *schemas*, rather than stages. Where Kohlberg considers developmental stages to be independent from philosophical distinctions, Rest finds such distinctions not particularly meaningful and argues that the consideration of an individual resolving a moral dilemma is the most useful analysis (Rest, 1979). Therefore, the terminology of moral schemas refers to general cognitive structures that provide a skeletal conception, exemplified (or instantiated) by cases or experience (Rest et al., 1997;1999).

Rest and colleagues (1999) describe moral judgment development according to three schemas: Personal Interest, Maintaining Norms and Postconventional. The definitions of these schemas are somewhat different from Kohlberg's stages, especially when looking at the distinction between conventional and post-conventional thinking. The schema Maintaining Norms, as derived from Kohlberg's stage 4, consists of five elements: need for norms, society wide scope, uniform, categorical application, partial reciprocity, and duty orientation. These elements connect law to order in a moral sense, leading to the expectation that without laws, there would be no order. The acquisition of this schema provides conventional thinkers with a sense of moral necessity for the maintenance of social order. In a rational reconstruction of the Postconventional schema, Rest proposes four elements: primacy of moral criteria, appeal to an idea, shareable ideals, and full reciprocity. The defining characteristic of postconventional thinking is that rights and duties are based on shared ideas for organizing cooperation in society. When comparing his theory to Kohlberg's, Rest and colleagues (1999, p. 40) said: "our notion of a schema is broader, less partisan and one could say it is more timid and less exact". By giving up the notion of defining 'hard' stages, Rest's theory envisions development as shifting distributions rather than a staircase. The more consolidated a person is in one of the schema's, the greater the ease and consistency in information processing. The post-conventional schema can be seen as a variant from the maintaining norms schema, as they are similar in the ease of information processing but differ in where they lead (for example: S56 favors the rights of homosexuals, S4 tends to do less).Therefore, it is common for a person to have a non-consistent view and mix arguments from both schema's, making it likely for them to be correlated (Rest et al., 2000).

Rest uses ranges of responses to represent the same types of reasoning in different manifestations, instead of considering and classifying every response separately, like Kohlberg. This led to a 'soft-stage concept', where the individual's moral reasoning level is represented by a composite of thinking in multiple, contiguous schemas. This differs from Kohlberg, who used discrete stage classifications with no stage mixtures, unless an individual is in a short 'transition phase'. Rest's model therefore suggests considerably different interpretation of moral reasoning levels, where the consideration brought up by is indicative for the developmental level (Elm & Weber, 1994; Rest 1979).

The Defining Issues Test (DIT)

Rest's well-known method for assessing moral judgment maturity is The Defining Issues Test (DIT) (Rest et al., 1999), derived from the time-consuming oral interview Kohlberg used. Whereas the MJI uses production tasks, by asking the participants to produce justifications, the DIT uses recognition tasks, where examples of ideas are provided to provoke a response. According to Rest, a production task tends to underestimate a person's understanding, which helps to explain why in Kohlberg's data postconventional stages were rare. The DIT ensures verbal expression to be less of a burden and reduces variability in interpretation of answers. The questionnaire was designed as a 'quick and dirty' alternative,

suitable for large-scale research. However, a disadvantage is that it is still lengthy and difficult to read for children and others with limited attentional and reading capacities, such as juvenile delinquents (Basinger & Gibbs, 1987).

Another Neo-Kohlbergian approach: Gibbs

Gibbs, who had worked in Kohlberg's research team for years, takes a somewhat different perspective on moral development. In his view, moral development concerns the development of perspective taking abilities in the moral realm. Without denying the importance of moral principles, moral reasoning development is fundamentally a process of perspective taking capacities. Therefore, Gibbs criticizes Kohlberg when it comes to using internalization as an explanation only (Gibbs & Schnell, 1985; Gibbs et al., 1992, Gibbs, 2003).

In Gibbs' revision, the lifespan development of moral reasoning includes two major phases: standard and existential. By terming these modes as 'phases', it is appropriate for them to overlap in time. The standard developmental phase consists of two overlapping levels: immature and mature, each containing two stages. The immature stages are constructed in childhood and represent relatively concrete and superficial thinking. In stage 1 (Centration), morality tends to be confused with physical size or power or with egocentric desires. The vulnerability of a young child to the immediately salient physical reality, is evident in moral, social, and non-social cognitive domains. Due to gains in mental coordination, logic-related inference and perspective-taking, stage 2 (Exchanges) brings out a more psychological and pragmatic morality. The mature stages are typically constructed during late childhood and adolescence, but there can be developmental delay, even in adults. The core appeal of stage 3 (Mutualities) is the third-person perspective, where an individual uses ideal reciprocity, mutual trust, and intimate sharing as the basis for interpersonal relationships. In stage 4 (Systems), moral maturity expands in terms of addressing the need for commonly accepted standards and values in complex social systems (Gibbs, 2013). The transitional phase 3-4 (Relativism of Personal Values) describes a form of moral reasoning that extends beyond interpersonal relationships but does not yet address functional requirements of society clearly (Gibbs et al., 1992). The postconventional level as described by Rest returns as the existential phase in Gibbs's theory. This phase exceeds the standard moral judgment stages and involves hypothetical contemplation and spiritual awakenings (Gibbs, 2019).

In contrast to Kohlberg and Rest, Gibbs argues that the postconventional level should not be regarded as the exclusive level of moral maturity. In fact, it should not even be a part of the standard stage sequence at all. In a critique (Gibbs, 1977; 1979), the postconventional level is said to be rather meta-ethical, than an expression of a more extensive structural development. Therefore, in Gibbs' theory (and instrument), the stages 3 and 4 already represent mature moral reasoning.

Sociomoral Reflection Form

      Because of the practical difficulties of the MJI, Gibbs initiated the development of a new production measure for moral reasoning: the Sociomoral Reflection Measure (SRM) (Gibbs et al., 1982). Within this measure, moral dilemmas play a significant role in the collection of moral judgment data. Dilemmas provide the subject with concrete situational details, without interference of preconceptions. The SRM allows group administration without individual probing, because of a questionnaire format with scorable responses. Based on the SRM, the Sociomoral Reflection Objective Measure (SROM) (Gibbs et al., 1984) was developed. Different than the SRM, the SROM asks subjects to select stage-significant reasons that are 'close' and 'closest' to the one they would find important in moral reasoning, and thereby provides an indirect index of sociomoral reflection. Due to a less complex scoring system than the MJI, the SRM and the SROM were a success (Gibbs et al., 1982; 1984). However, with an administration time comparable to the DIT (approximately 35-45 minutes), it still cannot be characterized as brief or simple. Then the thought occurred: were dilemmas really necessary? By exploring alternatives, Gibbs et al. (1992) discovered that a simple introductory phrase could provide enough adequate contextual stimulation to initiate reflection on moral issues, even in children and juvenile delinquents. They created the Sociomoral Reflection Measure – Short Form (SRM-SF. Gibbs et al. (1992) believed that the SRM-SF is a highly successful and practical production measure of moral reasoning. The SRM-SF laid the basis for the SRM-SFO, a recognition task. The SRM-SFO is a questionnaire without the use of moral dilemmas. Being half the length of the SROM because it contained no dilemmas and entailing a simpler question format than the SRM-SF, because respondents react to stage-typed reasons, the SRM-SFO could become a successful measure for moral reasoning in large scale research (Brugman et al., 2021).

Purpose of the present study

      To date, to my knowledge, there is no research available where both the DIT and the SRM-SFO are used together to assess moral judgment maturity. This is partly due to the different age groups the measures are targeted to: adolescents and (young) adults, particularly college students, by the DIT, and adolescents, particularly juvenile delinquents, by the SRM-SFO. However, the SRM-SFO has recently also successfully been used in college students (Shields et al., 2018), making a comparison between both instruments meaningful in the age group 16 to 21 years. This research therefore focuses on the conceptual and instrumental similarities and differences in approach and measurement of moral maturity, according to Rest and Gibbs. The main differences between both neo-Kohlbergian approaches are listed in Table 1.

Table 1

Most important differences between the neo-Kohlbergian theories and instruments of Rest and Gibbs.

Rest	Gibbs
Theory	
- Moral judgment develops through rights and responsibilities in a social system.	- Moral judgment develops based on the development of perspective taking abilities.
- Moral maturity is represented by the post-conventional stages 5 and 6 by Kohlberg.	- The mature level is reached in the conventional stages (3 and 4 by Kohlberg), the post-conventional stages are meta-ethical.
Measurement	
- The consideration brought up by an individual resolving a moral dilemma is the most useful analysis	- A simple introductory phrase provides enough adequate contextual stimulation to initiate reflection on moral issues
- Recognition measure (DIT)	- Production (SRM-SF) /recognition measure (SRM-SFO)

When assessing moral judgment maturity, certain relationships are expected with demographic characteristics (Stams et al., 2006). Since moral judgment develops over time, age is one of the most important variables to consider. Second, when looking at gender, females have been found to reach higher stages of moral judgment sooner than males, when transitioning to adolescence (Garmon et al., 1996). Third, following the cognitive-developmental approach, higher-level education reflects a higher capacity for abstract thinking. Therefore, a higher intelligence is also related to more advanced stages of moral reasoning (Colby et al., 1983). Previous studies have also found lower levels of moral maturity in delinquent adults, that generally show higher levels of anti-social behavior (Blasi, 1980; Nelson et al., 1990; Smetana, 1990). The present study expects the DIT and SRM-SFO to find to be consistent in establishing similar patterns with these demographic characteristics and (self-reported) anti-social behavior. This led to the formulation of the following two hypotheses:

1. The DIT and the SRM-SFO show considerable consistency in predicting moral judgment maturity in relation to demographics such as age, gender, and educational level.
2. When using self-reported anti-social behavior as a referent, both the DIT and the SRM-SFO predict lower moral judgment maturity scores when a person scores higher on anti-social behavior.

Further, given that a moral maturity score on the DIT is based on a level, that does not even exist in Gibbs' theory, while the moral maturity score on the SRM-SFO is based on the stages that are neglected when using the DIT, one wonders how these different measures of moral maturity succeed in establishing similar empirical relationships and how they are related to each other. The moral maturity score on -the easier to use- SRM-SFO, based on Gibbs' standard stages three and four is expected to be close enough to the moral maturity score on the DIT, based on Rests' post-conventional level, to be used as alternative. This led to the formulation of a third hypothesis:

3. When comparing the moral maturity score on the well-known DIT (P-score and alternative indices) to the moral maturity score on the SRM-SFO (SRMP-score), a positive correlation is expected.

Method

Sample and participant selection

Data was collected using an online survey in the Netherlands. The total sample consisted of 145 respondents (30% males, 70% females) between 16 and 21 years old, heterogenous to educational level. Many participants (54%) were excluded from further analyses because of missing values or inconsistencies. This resulted in a final sample of 79 participants (32% males, 68% females), with an average age of 19.7 years ($SD = 1.26$). A distinction was held between lower educational levels (practical and (pre-)vocational secondary education; 8%), middle educational levels (senior general secondary and higher vocational education; 60%) and higher university educational levels (33%).

Procedures

Measures were administered using an online survey tool (Qualtrics). Students at the University of Utrecht were granted study points (PPU) for participating, other students and juveniles were gathered at different educational institutions and participated voluntary. All participants gave consent for participation and were informed about the goals of the study. The research was approved by the Ethics Review Board of the Faculty of Social & Behavioral Sciences of the University of Utrecht.

Measures

DIT-2

The original DIT2 (Rest et al., 1999) consists of six stories and takes about 45 minutes to complete. To reduce the completion time of the total survey, the short form of the DIT including three stories, was used in the present study. In each set the respondent is presented with a moral dilemma (e.g., the story of a man stealing food for his starving family), based on the dilemmas used by Kohlberg. Next

are 12 items presenting an issue for consideration in solving the dilemma (e.g., 'steal the food', 'do not steal' or 'can't decide'). The respondent's must rate each statement according to its importance in making a moral decision ('great', 'much', 'some', 'little', or 'no'). These statements represent the stages of Rest in a random order, together with a view 'meaningless' and 'anti-social' statements. After rating the 12 items, the participant is asked to reconsider all items simultaneously and to rank the four most important items in making a decision ('most important', 'second most important', 'third most important', 'fourth most important').

The ratings and rankings are used to derive a respondent's score on the DIT2. The weighted sum of ranks for the postconventional items, form the P-score. For instance, when a post-conventional item is ranked as 'most important', the P-score increases with four points. Ranking this item as 'second most important', would increase the P-score with three points. Important to consider is that the number of P items is not four in every story. Therefore, the total score ranges from 0 to 25 in the three-story version. Raw P scores are converted into percentages, interpreted as the degree to which the participant thinks postconventional considerations are important. Rank-rate inconsistencies are assumed to indicate random answers. Respondents who have more than two inconsistencies in a story or more than one story with inconsistencies, are declared missing. This is also the case for respondents who gave more than four 'meaningless' answers.

SRM-SFO

The SRM-SFO (Brugman et al., 2021) consists of ten sets of questions. Each set introduces a value, represented by an everyday context. These values have been organized in four areas: Contract and Truth (item 1-4), Affiliation (item 5), Life (item 6-7), and Legal Justice (item 8-10). The lead-in of a set is a simple statement such as: "Think about when you've made a promise to a friend of yours." Next, the respondents are asked evaluation questions (e.g., "How important is it for people to keep promises, if they can, to friends?"). They must consider whether the represented value is 'very important', 'important' or 'not important' to them. Subsequently, questions for the reasoning or justification for support of the importance of that value are asked. These questions are generic stage-related statements, representing the Gibbs' stages 1-4 in a random order. Respondents rate each statement on three alternatives: capturing or being close to their own justification, not being close to their justification, or not sure with regard to their own justification. Finally, the respondent selects the statement that is the closest or most representative to their own justification.

Each set has a maximum of two close mature responses and one closest mature response. The percentage of the mature responses accepted by the respondent, from the total number of potential mature responses, is called the Sociomoral Reflection Maturity Percentage (SRMP). The SRMP can vary between 0 (completely immature) to 100 (completely mature). The consistency check declares that a closest mature response is only valid if the respondent accepted the corresponding close mature score.

A respondent's score is declared missing when a close, closest or item score is missing or when less than seven items were completed.

ABQ

The Antisocial Behavior Questionnaire (ABQ) (Leenders & Brugman, 2005) is a self-report measure for assessing delinquency and anti-social conduct. The ABQ consists of 12 questions, that refer to theft (4 items), robbery (2 items), physical aggression toward others (3 items), and vandalism (3 items). Behaviors were rated on a 4-point scale (never, once, sometimes, often), with a total score of maximum 48 points. The reliability of the twelve items scale was .614 (Cronbach's Alpha).

Analyses

Descriptive statistics were presented for the demographic variables (age, gender, educational level), antisocial behavior, and moral maturity together with a normality check including assessment of skewness and kurtosis. Moral maturity scores on the DIT were divided into different P-scores: a post-conventional score (loading on stages 5 and 6 items), a maintaining norms score (loading on stage 4 items), a personal interest score (loading on stages 2 and 3 items), and a combined score (loading on stages 3, 4, 5 and 6 items). Reliability of both instruments was assessed thoroughly, because of the importance in the context of answering the main question. Validity was assessed using correlational techniques focused on the expected patterns between the DIT/SRM-SFO variables and other variables (age, gender, educational level, and anti-social behavior), and to similarities and differences of moral maturity, as defined by Rest and Gibbs. Further assessment of significant relationships has been done by the help of T-tests and regression analyses. Pearson correlation coefficients of maturity scores on both measures were calculated, where different indeces of the DIT were compared to the SRMP scores of the SRM-SFO. Analyses were executed in SPSS28.

Results

Descriptive analysis

In the community sample, the distribution of moral maturity scores on the SRM-SFO was negatively skewed to the left (skewness = -2.1, kurtosis = 5.7). The kurtosis may indicate that the values are not normally distributed. Looking at item level, the skewness and kurtosis scores appear to be reasonable, with some outliers for the kurtosis. This could potentially be explained by the fact that the sample mainly consists of higher educated people. The skewness and kurtosis of the different P-scores on the DIT were all between -2 and 2 and therefore considered normally distributed.

The reliability of items contributing to the moral maturity score on the SRM-SFO appeared to be acceptable, as shown in Table 1. The reliability analysis of the DIT was more complicated, as

respondents had to fill in both ratings and rankings. Reliability of rating scores for different the P-scores were first assessed, as shown in Table 2. Reliability of the personal interest (P23) and maintaining norms (P4) rating scores appeared to be unacceptable low. The reliability of the post-conventional rating scores (P56) was doubtful, while the reliability of the combined P-score ratings (P3456) was acceptable for further research. Reliability of the ranking scores only have also been assessed, as shown in Table 2.

Table 2

Reliability analyses of P-scores (DIT) and the SRMP-score (SRM-SFO).

	P23	P4	P56	P3456	SRMP
Cronbach's α	.40	.44	.61	.67	.68

Note. α > .60 was considered acceptable for research.

Hypothesis 1 and 2

An overview of the mean scores on the SRM-SFO, DIT and ABQ is provided in Table 3. According to the SRMP score of the SRM-SFO, the total sample on average used forms of moral reasoning on a mature level. When looking at the combined P-score (P3456) as the moral mature stages of the DIT, the total sample on average mostly used mature forms of moral reasoning.

No significant relationships were found between the three demographic variables. Therefore, partial correlations do not have to be taken in account. When looking at gender differences, females scored on average higher than males on moral maturity on both the SRM-SFO and the DIT. This difference for gender was significant for the both the SRMP ($t(77) = -2.19$, $p < .001$) and the combined P-score ($t(77) = -0.86$, $p < .05$). Males reported slightly higher on anti-social behavior than females, but this difference was not significant.

Second, when looking at age subgroups, higher average moral maturity scores on the SRM-SFO were found in higher age categories. There appeared to be no clear relationship with age and moral maturity scores on the DIT. Younger respondents on average presented a higher score on anti-social behavior than older respondents.

Finally, when looking at education, higher levels of education on average scored higher on moral maturity on the DIT. A significant difference was found between the lower and higher educational level ($t(8.3) = -2.1$; $p < .05$). There appeared to be no clear relationship with educational level and moral maturity scores on the SRM-SFO. The lower educational levels on average presented a lower score on anti-social behavior than higher educational levels, but these differences were not significant.

Table 3

Mean percentage scores on the SRM-SFO and DIT and mean scores on the ABQ (M (SD)).

	n	SRMP	P23	P4	P56	P3456	ABQ
Total sample	79	81.0 (16.12)	36.7 (15.28)	24.1 (12.93)	30.7 (14.09)	80.9 (14.45)	15 (3.16)
Gender							
Male	25	73.4 (24.60)	44.2 (15.92)	21.8 (14.77)	24.2 (12.39)	78.9 (17.86)	16 (3.86)
Female	54	84.3 (10.03)	33.2 (13.83)	25.3 (11.97)	33.7 (13.90)	81.9 (12.65)	15 (2.76)
Age							
16/17	5	69.3 (31.80)	37.2 (17.15)	18.0 (11.37)	33.4 (15.07)	84.2 (14.90)	19 (7.14)
18/19	25	78.0 (20.24)	36.7 (17.84)	22.8 (12.02)	31.7 (16.02)	77.5 (15.41)	15 (2.90)
20/21	49	83.5 (11.22)	36.0 (13.51)	25.7 (13.54)	30.3 (13.13)	82.1 (14.03)	15 (2.48)
Educational level							
Lower	6	81.0 (14.15)	43.3 (24.44)	22.2 (11.80)	24.0 (20.00)	74.0 (22.52)	13 (2.07)
Middle	47	79.0 (19.43)	37.7 (15.34)	23.2 (13.64)	29.5 (14.30)	79.3 (15.43)	15 (3.52)
High	26	84.0 (8.88)	33.3 (12.30)	26.4 (12.00)	34.4 (11.63)	85.5 (8.64)	15 (2.62)

Note. The SRMP reflects the moral mature score on the SRM-SFO. For the DIT, P23 represents the Personal Interest Schema, P4 the Maintaining Norms schema, P56 the Post-conventional schema and P3456 the combined P-score. ABQ stands for the self-reported anti-social behavior.

Convergent validity of moral maturity scores on the SRM-SFO was assessed through correlations with the theoretically relevant variables, as shown in Table 4. The SRM-SFO demonstrated good convergent validity through significant correlations with gender ($r = .32$), age ($r = .21$), and anti-social behavior ($r = -.27$). No significant correlation was found between moral maturity scores and educational level, also with age partialled out. Considering the non-heterogenous sample of this sample, the instrument could be sensitive in investigating developmental age trends in a more homogenous sample.

The DIT demonstrated poor convergent validity, looking at the establishing of relationships with theoretically relevant variables. In line with the expected, significant correlations with gender were found in the P56 score ($r = .316$), and the P23 score ($r = -.336$). The combined P-score succeeded in predicting a correlation with educational level ($r = .240$), where other P-scores did not, also with age partialled out. No effects were found between P-scores and anti-social behavior. Compared to the SRM-SFO, the DIT did not show as much consistency in predicting moral judgment maturity in relation to relevant variables.

In a stepwise regression with self-reported anti-social behavior as dependent variable, and the demographic variables and the SRMP score as explanatory variables, only the instrument revealed a

significant relationship ($F(-5.25) = 5.95, p < .05$). The same analysis with the different P-scores revealed no significant relationships.

Table 4

Correlations of moral maturity scores (SRM-SFO/DIT) with demographics and anti-social behavior.

	Gender	Age (in years)	Educational level	ASB
SRMP	.315**	.284*	.114	-.271*
P23	-.336*	.016	-.128	.007
P4	-.129	.055	.117	-.148
P56	.316**	-.009	.210	-.051
P3456	-.089	.088	.240*	-.148

*Note. **p <.001, *p < .05*

Hypothesis 3

An overview of the correlation coefficients between P-scores and the SRMP is presented in table 4. First, different P-scores were compared to each other, where the lowest P-score (P23) seemed to correlate progressively more negative to higher P-scores. This is in line with the idea that a person who prefers more immature reasonings, scores lower on stages that contain more mature reasonings and vice versa.

Second, the moral maturity score of the SRM-SFO was compared to the moral maturity scores of the DIT. The original P-score 'maintaining norms' (P56) correlated significantly to the SRMP ($r = .329$), which is in line with the hypothesis that the moral mature stages of Rest are linked to the moral mature level of Gibbs. The combined P-score correlated even better to the SRMP score ($r = .415$), at a significant level. This correlation coefficient between the combined P-score and SRMP becomes slightly higher when age is partialled out ($r = .424$).

Table 5

Pearson correlations between different P-scores and the SRMP.

	P23	P34	P4	P56	P3456
P23					
P34	.353**				
P4	-.448**	.489**			
P56	-.558**	-.551**	-.207		
P3456	-.149	.581**	.342**	.359**	
SRMP	-.139	.085	.131	.329**	.415**

*Note. **p <.001, *p < .05*

Discussion

The purpose of this study was to gain a better understanding of the measuring of moral judgment maturity, from a cognitive developmental point of view. Two Neo-Kohlbergian approaches were explained and compared to each other: the soft-schema model by Rest and the phase-model by Gibbs. From these theories, two different instruments (respectively the Defining Issues Test and the Sociomoral Reflection Measure) have been created with the same aim: to measure moral reasoning development by a questionnaire. In this study, an answer was sought to the question how these instruments succeed in establishing similar empirical relationships and how they are related to each other.

The present study showed three key findings in line with our hypotheses. First, the results showed several patterns between moral maturity scores of both measures with demographic variables, consistent with what was expected from the previous literature (Colby et al., 1983; Garmon et al., 1996; Stams et al., 2006). The SRM-SFO showed the expected effects of gender and age. The DIT showed less consistency and demonstrated only small effects of gender and educational level in higher P-scores. Second, when using self-reported anti-social behavior as a referent, only the SRM-SFO succeeded in presenting a relationship consistent the previous literature (Mullis et al., 2004). No relation between self-reported anti-social behavior and the DIT was found. These first two finding are in favor of the SRM-SFO.

Finally, significant relationships between the moral maturity scores of the SRM-SFO and the DIT (P56) were found. This finding can be explained by the idea that the moral mature schemas of Rest are linked closely to the moral mature level of Gibbs. This relationship becomes stronger when the schema's 3 and 4 by Rest are added in the moral maturity score of the DIT (the combined P-score). Because in Gibbs' point of view, moral maturity is reached already in the stages 3 and 4, it can be explained that adding these stages to the moral maturity score by of Rest leads to a higher correlation. This finding is therefore consistent with the literature, keeping in mind that Kohlberg's stages form the base of both theories. Whether this effect reflects a difference in formulations of the stages of both theories or an instrument effect called 'cherry picking' needs further investigation. For follow-up research it could therefore be important to control for social desirability.

This study has several limitations. One limitation is that many participants (53%) had to be excluded because of missing data and inconsistencies, which led to a rather homogenous sample of higher educated and older participants. It also turned out to be more difficult than estimated in advance to reach that younger and lower-educated group, making the data skewed to the right. Perhaps, stronger relations between moral judgment maturity and age or educational level could have been found with a more heterogenous sample of participants.

It is possible that a large part of the drop-out is due to the total length of the online survey. According to a study about the effect of questionnaire length on response quality (Galesic & Bosnjak, 2009), an online survey longer than 10 minutes already shows reduced completion rates. In this research

the shortened version of the DIT was used, with the intention to reduce the extensiveness of the total survey. However, the average time of the survey was still about 28 minutes. Due to limited time, one survey including the three questionnaires was used. However, for future comparative research on the DIT and SRM-SFO it is advised to do it in two parts. Physical classroom administrations could also provide more reliable data, but this was no option in times of pandemic.

Another important limitation to address is the reliability of both instruments. Using a shorter version of the DIT generally lowers the reliability and correlation with external variables by about ten points (Rest et al., 1977). To minimize this trade-off, the recommended combination of stories is used. Yet, values of Cronbach's alpha still varied from unacceptable to just acceptable for research. The reliability of the SRM-SFO was also just acceptable enough. This is also a reason to interpret the results with a little more caution. One last point concerns the representativity of the SRM-SFO, that only contains items that represent the standard phase. By adding items of the existential phase, stages 5 and 6 of Rest's model are better covered. This would have made a comparison between both measures even more meaningful.

Despite these limitations, this research enhanced the understanding of the relationship between the theories and measurements of Rest and Gibbs, that have not been directly linked yet. It is a first step in integrating two lines of research and it provides us with reasons to believe that the -easier to use- SRM-SFO is the better alternative in measuring moral judgment maturity in group of 16- to 21-year-olds.

References

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

Brugman, D., Beerthuizen, M. G., Helmond, P., Basinger, K. S., & Gibbs, J. C. (2021). Assessing moral judgment maturity using the Sociomoral Reflection Measure—Short Form Objective. *European Journal of Psychological Assessment*.

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

Blasi, A. (1990). How should psychologists define morality? or, the negative side effects of philosophy's influence on psychology. In Wren, T.E. (Ed.), *The Moral Domain*: Essays in the Ongoing Discussion Between Philosophy and the Social Sciences. MIT Press.

Carpendale, J. I. (2000). Kohlberg and Piaget on stages and moral reasoning. *Developmental Review*, *20*(2), 181-205.

Colby, A., Kohlberg, L., Gibbs, J., Lieberman, M., Fischer, K., & Saltzstein, H. D. (1983). A longitudinal study of moral judgment. *Monographs of the society for research in child development*, 1-124.

Colby, A. & Kohlberg, L. (1987). *The measurement of moral judgment: Theoretical foundations and research validation*. Cambridge University Press.

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public opinion quarterly*, *73*(2), 349-360.

Garmon, L. C., Basinger, K. S., Gregg, V. R., & Gibbs, J. C. (1996). Gender differences in stage and expression of moral judgment. *Merrill-Palmer Quarterly (1982-)*, 418-437.

Gibbs, J. C., & Schnell, S. V. (1985). Moral development" versus" socialization: A critique. *American Psychologist*, *40*(10), 1071.

Gibbs, J. C., Widaman, K. F., & Colby, A. (1982). Construction and validation of a simplified, group-administerable equivalent to the moral judgment interview. *Child development*, 895-910.

Gibbs, J. C., Basinger, K. S., & Fuller, D. (1992). *Moral maturity: measuring the development of sociomoral reflection*. L. Erlbaum Associates.

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

Gibbs, J. C. (1979). Kohlberg's moral stage theory. *Human development*, *22*(2), 89-112. Kohlberg, L., & Kramer, R. (1969). Continuities and discontinuities in childhood and adult moral development. *Human development*, *12*(2), 93-120.

Gibbs, J. C. (2019). *Moral development & reality: Beyond the theories of Kohlberg, Hoffman, and Haidt* (4th Ed.). New York: Oxford University Press.

Kohlberg, L., Boyd, D.R., & Levine, C. (1990). The return of stage 6: Its principle and moral point of view. In Wren, T.E. (Ed.), The Moral Domain: Essays in the Ongoing Discussion Between Philosophy and the Social Sciences. MIT Press.

Kohlberg, L. (1984). *Essays on moral development: Vol.2. The psychology of moral development.* San Francisco: Harper & Row.

Leenders, I., & Brugman, D. (2005). Moral/non-moral domain shift in young adolescents in relation to delinquent behaviour. *British Journal of Developmental Psychology*, *23*(1), 65–79.

Maxim, S. T. (2014). Ethics: Philosophy or science? *Procedia-Social and Behavioral Sciences*, *149*, 553-557.

Mullis, R. L., Cornille, T. A., Mullis, A. K., & Huber, J. (2004). Female juvenile offending: A review of characteristics and contexts. *Journal of Child and Family Studies*, *13*(2), 205–218.

Nelson, JR, Smith, JJ and Dodd, J. 1990. The moral reasoning of juvenile delinquents: a meta-analysis, *Journal of Abnormal Child Psychology*, 18: 231–239.

Piaget, J. (1960). *The general problems of the psychobiological development of the child.*

Piaget, J. (1965). *The moral judgment of the child.* New York: The Free Press.

Rawls, J. 1971. A Theory of Justice. *Ethics*, 229-234

Rest, J., Thoma, S. J., Narvaez, D., & Bebeau, M. J. (1997). Alchemy and beyond: Indexing the Defining
Issues Test. *Journal of Educational Psychology, 89* (3), 498-507.

Rest, J., Narvàez, D., Thoma, S., & Bebeau, M. (1999). DIT2: Devising and testing a revised
instrument of moral judgment. *Journal of Educational Psychology*, *91*, 644–659.

Rest, J. R., Thoma, S. J., & Bebeau, M. J. (1999). *Postconventional moral thinking: A neo-Kohlbergian
approach*. Psychology Press.

Rest, J. (1999). *Postconventional moral thinking: a neo-Kohlbergian approach*. Lawrence
Erlbaum Associates.

Rest, J. R., Narvaez, D., Thoma, S. J., & Bebeau, M. J. (2000). A neo-Kohlbergian approach to morality
research. *Journal of moral education*, *29*(4), 381-395.

Shields, D. L., Funk, C. D., & Bredemeier, B. L. (2018). Relationships among moral and
contesting variables and prosocial and antisocial behavior in sport. *Journal of Moral Education, 47*, 17
33. https://doi.org/10.1080/03057240.2017.1350149

Smetana, J. G. (1990). Morality and conduct disorders. In *Handbook of developmental
psychopathology* (pp. 157-179). Springer, Boston, MA.

Stams, G. J., Brugman, D., Deković, M., Van Rosmalen, L., Van Der Laan, P., & Gibbs, J. C.
(2006). The moral judgment of juvenile delinquents: A meta-analysis. *Journal of abnormal child
psychology*, *34*(5), 692-708.