Measuring Moral Judgments from a Neo-Kohlbergian Perspective.

Steve Thoma

These are unpublished conference papers for the 'Can Virtue Be Measured?, held by the Jubilee Centre for Character and Values at Oriel College, Thursday 9th – Saturday 11th January 2014. These papers are works in progress and should not be cited without author's prior permission.

School of Education

University of Birmingham

Edgbaston

Birmingham

Measuring Moral Judgments from a Neo-Kohlbergian Perspective.

(Draft)

Steve Thoma

The University of Alabama

During the late 1990s, four of us met weekly to take stock of our program of research. The group consisted of Jim Rest a student of Larry Kohlberg and the developer of the Defining Issues Test of moral judgment development; Mickey Bebeau, a colleague from the school of health ecology who has a long standing interest in professional ethics education; Darcia Narvaez who brought an interest in schema theory and moral text processing to the group and myself with my interest in measurement and personality development. We spent a lot of time discussing the strengths and weaknesses of our research program and soon decided to capture these discussions in a book which we entitled Postconventional Moral Thinking a Neo-Kohlbergian Approach. This book laid out what we learned from our work using the DIT and which aspects of our theoretical position we could maintain and where modifications were required. This talk will review the positions we took in framing the Minnesota approach and then identify research traditions that have been particularly influenced by our model.

Why identify our position with Kohlberg?

Kohlberg's theory dominated the field for decades and his legacy is still evident in more current models of moral functioning. Of particular importance were his efforts to bring a Piagetian perspective on moral development to America and then fusing these Piagetian notions with the then current work by the philosopher: John Rawls. In so doing Kohlberg provided a new way of understanding human cooperation. During the 60s and 70s, Kohlberg's theory and measurement system were the gold standard by which all other models of moral thinking were compared. Furthermore, Kohlberg's theory had a lot to say about the social issues of the day including the Vietnam War, the civil rights movement and the political turmoil of the 60s.

It was during this heady time in moral psychology that Rest developed the DIT. It grew out of Kohlberg's work and Rest had a long association with the Harvard group and Kohlberg in particular. Indeed, the DIT gained early acceptance by its close relationship to the Kohlberg

approach. It is important to note that all of the stories used by the original DIT originated from Kohlberg's work. Similarly, the items used on the DIT were distillations of participant interview responses on the Kohlberg interview.

At first, the view in the field was that the Minnesota group had developed a "quick and dirty" objective measure of Kohlberg's theory that might have its uses when one could not use the preferred Kohlberg interview approach. However over time and with the accumulation of data sets numbering in the hundreds of thousands, the two research programs began to diverge on key theoretical issues. Reflecting these divisions across the 80s and 90s, the Minnesota approach became viewed as a distinct theoretical branch of the cognitive developmental approach to moral psychology. In the late 1990s, as our group reviewed the Minnesota position, some of us pushed the group to distance our approach from our Kohlbergian roots but Rest was committed to the notion that we respect and acknowledge our connections to Kohlberg and to reaffirm some basic points of agreement. To that end, we began our deliberations looking for points of overlap with the Kohlbergian model.

Similarities with Kohlberg's theory.

The basic framework of the Neo-Kohlberian model significantly borrows from Kohlberg's approach. Central to both perspectives is a focus on cognition. Like Kohlberg we reaffirmed the notion that through interactions within the social world the individual comes to develop an understanding of social cooperation—what is owed and what one owes others. This view holds that a model of moral functioning must include a cognitive focus as it is essential for a complete understanding of moral functioning more broadly. Secondly, and consistent with the Piagetian perspective, we agree with Kohlberg that the individual does not passively accumulate information about the social world such as an understanding of rights, duties, and ideological positions. Instead, social information is self-constructed and organized by the individual. Thirdly, we agree that the understanding of social moral concepts is developmental and can be viewed as moving from less complex and incomplete understandings to more defensible and elaborate positions. That is, differences between people in their moral understanding can be explained in part by a developmental dimension that reflects the complexity of the ideas the individual uses to interpret moral phenomena. Finally, across individuals the central

developmental feature that defines the second decade of life and beyond is the transition from a conventional perspective to a post-conventional understanding of cooperation. In this view, adolescence is the time in which one comes to understand the moral basis of convention and how these normative systems work to regulate society. In late adolescence and into the adult years, this conventional view is supplanted by a growing awareness of the role that to be moral, conventions must conform to a shared ideal notion of cooperation.

In addition to these four main theoretical assumptions, the neo-Kohlbergian view makes a distinction between two views of moral functioning often conflated in the application of Kohlberg's theory. We note that it is helpful to distinguish a focus on moral thinking as it applies to society-wide social structures from the interpersonal morality of everyday life. At the society-wide or macro- morality level Kohlberg's theory is applied to the individual's understanding of the moral basis of laws, governing structures and general practices of society. In this Macro-morality perspective, the idealized perspective is one that prioritizes principle over partisanship and impartiality over favoritism. In contrast, the morality of everyday life, or micro-morality attends to the understanding of how morality underlies human exchanges including being empathetic, kind, and courteous. The micro morality focus is on how the individual emphasizes positive interpersonal characteristics in interactions with special individuals. Clearly there are both tensions and communalities between these conceptions. We have argued, however that Kohlberg's theory is a better description of macro-moral thinking and we focus on macro-moral thinking in our measurements.

Differences from Kohlberg's theory

The points of convergence with Kohlberg's theory notwithstanding, the neo-Kohlbergian model departs from Kohlberg's stages and sequences in significant ways. Two early modifications were particularly noteworthy and signaled that the Minnesota group's perspective was increasingly distinct from Kohlberg's views. First, was the move to abandon the orthodox stage model that was so much of a feature of the Kohlberg approach. In the 1970s, Rest questioned Kohlberg's acceptance of a strong stage model of development in which individuals move from stage to stage one stage at a time. Instead, the DIT supported a developmental model that defines growth as a gradual shift from lower to more complex conceptions of social/moral

cooperation. Furthermore, DIT researchers assume that at any given time there are multiple conceptions available to the individual. Thus, appropriate measurement strategies must assess not only which conceptions are available, but also the most preferred system.

The second modification occurred in the 1980s and was the result of a growing perspective within the Minnesota group that a singular focus on moral judgment development was leading to diminishing returns. The future, it seemed, was linked to an expanded view of the processes involved in moral functioning and the need to move away from a global stage model. The resulting Four-Component model (FCM) became the framework for the work that followed into the 90s and beyond. According to the FCM, moral actions are the result of a least four component processes operating individually and in interaction. These component processes that describe the moral system include processes that promote an individual's ability to identify and attend to moral issues (i.e., moral sensitivity), the ability to reason and justify the morally ideal course of action (i.e., moral judgments); a motivational system that prioritizes the morally ideal choice against other claims on the individual (i.e., moral motivation), and finally, a system that can construct an appropriate action and stay on task (i.e., behavioral construction or moral character). It is important to note that the FCM regulates moral judgments as measured by the DIT to Component 2. Furthermore, it is only one of many strategies an individual can use to construct an idealized response to a moral event. In addition to moral judgment processes, one might prioritize social norms, religious prescription, among others. In addition, Rest and others were quick to note that the four components contained affective as well as cognitive processes and operated in a highly interactive way. That is, there was no a priori reason to expect moral actions to be the result of a simple linear sequence starting at component 1 and moving linearly to component 4.

These two changes set the stage for the more significant modifications which occurred in the 1990s. Chief among these changes was the adoption of a schema view of moral judgment development. The use of schemas to define our model of moral judgment development highlighted the significant differences in the focus of our model and assessment process. To contrast Kohlberg's view with our own, Kohlberg's moral stages are described in terms of cognitive operations that directly describe the structure of moral thinking. That is, the Kohlberg scoring system has been designed to separate the content of one's thinking from the cognitive

structure defined by stages. Thus, when discussing a moral dilemma, the situation, protagonist roles, etc. are content and not central to the assessment process. In so doing, the researcher assumes that he/she is able to directly measure the cognitive operations the individual uses to make judgments about moral content freed of the content itself. By contrast, we suggest that the attempt to focus on structure is problematic because there is little evidence that verbal utterances accurately capture the processes that structure our thinking. In our view, a more appropriate conception of what develops in moral thinking is a schema view that highlights how individuals understand, organize and prioritize moral content such as societal norms, systems and organizations. I will say more about schemas later on in the talk.

Consistent with our position on Kohlberg's assessment process and his claims about directly accessing cognitive operations is a more specific concern over the priority he placed on verbal data. It is common to read how verbal data is the preferred and most reliable means of assessing moral judgments. In this view, when participants are asked to explain their moral judgments the resulting information is particularly valuable in isolating the psychological processes that inform these judgments. We note that in the field of cognitive science there is little support for the view that participants have any insight into the processes that led to their judgments. Indeed, there seems little to support the privileged standing of interview over recognition data. A case in point is the relatively rare occurrence of post conventional moral thinking using Kohlberg's interview process compared to the more frequent rates described by the DIT. Individuals may be able to recognize the superiority of postconventional strategies without being able to articulate and defend these strategies. Our view is that these is a utility in focusing on tacit knowledge as it may be more influential in the resulting decisions and resulting moral actions.

The more relaxed view of content and structure we proposed further influenced our view on Kohlberg's universality claim. Kohlberg was clear that his stages were by definition universal because his model was based on cognitive operations that in turn, were linked to social and cognitive development models that were in turn assumed to be universal. Additionally, Kohlberg argued that a universality claim was essential to avoid moral relativism, which might allow for communities to define the moral system in any way they wanted (e.g., cannibalism is fine for cannibals). Our shift toward a model that allows for both content and structure in the descriptions of moral judgment development makes a universality claim more difficult since

individuals living in different communities—let alone different cultures-- will experience a range of social roles, norms and organizations and thus may have different ways of structuring moral content to derive moral judgments. The solution we offered reduces the universality claim to an empirical question. We presuppose that different communities have a mix of common and unique experiences that frame the social construction of a moral perspective at any given time. We further suggest that these different histories, institutional arraignment, and current concerns, are debated within the community and become shared experiences that inform individual moral thinking. This view equates morality with common law which also shares some common principles across cultures but also some unique features based on the specific experiences of the various communities. Taken together, we assume that there is some movement in the common morality over time and as with other systems such as law, and science, common morality evolves as new precedents and data are assimilated. We recognize the problem that moral relativism creates, however we would argue that a common morality developed through open discussion and applying shared ideals, is a substantial buffer to the worst excesses of relativism that Kohlberg so feared. Empirically, therefore, we make the case that universality is an empirical question that ought to be tested in order to identify common and unique elements across communities and cultures.

How is development defined in the Neo-Kohlbergian model?

In addition to altering the developmental model underlying the measure, w also have focused on how best to define the developmental dimension measured by the DIT. In its original conception, the DIT assessed a developmental dimension defined in terms of Kohlberg's stages as they were described in the early 70s. More recently, however, the fit of Kohlberg's model to DIT data has been assessed. Based on empirical studies using large and diverse samples including some with as many as 44,000 participants, the description of what the DIT measures have changed. Specifically, empirical estimates of the ways in which DIT items cluster suggest that the six stages described by Kohlberg do not fit the data. Instead, the obtained number of item clusters suggests three distinct groupings: Stage 2 and 3, Stage 4, and Stage 5 and 6. The finding of three distinct clusters is especially clear when the assessment is based on a heterogeneous sample including participants ranging from high school through the adult years. That is, empirically, the best fitting scheme based on DIT data is no longer the six Kohlberg

stages. Instead a three level model loosely informed by Kohlberg's model seems more appropriate.

It seems plausible that the obtained clusters are due in part to the adolescent and adult populations typically studied by DIT researchers and perhaps the properties of the DIT itself. However, empirically, it seems clear that participants taking the DIT tend to view items representing Stages 2 and 3 as less important reasoning than items in other clusters. Taken together, the stage 2 and 3 items are not often ranked; although attraction to these items is growing. That is, items that highlight self-preservation, self-interest, and personal relationships are viewed together as personal concerns that are not as central as other more-system wide issues represented by the stage 4 items and those that form the post conventional cluster. Unlike the stage 2 and 3 cluster, the stage 4 and postconventional items are often ranked and viewed as highly important. These findings support the view that the DIT items are assessing moral judgment development at the macro-moral level since the power of the DIT derives from the Stage 4 conventional items and the post-conventional items.

Interpreting the three clusters of items.

The three clusters of items suggest that the DIT measures three distinct moral schemas that are developmentally ordered. These schema are labeled: the Personal Interests schema (combining elements of Kohlberg's descriptions of Stages 2 and 3); the Maintaining Norms schema (derived from Kohlberg's definition of Stage 4); and the Post-conventional schema (drawing from Kohlberg's Stages 5 and 6—and equivalent to the items forming the original summary index called the P score). A description of each schema is presented below.

Personal Interest schema. We describe the main focus of the personal interest schema as highlighting a perspective that attends the gains and losses each individual may personally experience within a moral dilemma. Similarly, no attention is given to the larger social systems within this schema. Overall, as viewed through a personal interest lens, the social world is a loosely tied network of micro-moral considerations linking close relationships and individual interests. The Personal Interest Schema is fully developed by the time participants are able to reliably complete the DIT (typically defined as a 9th grade reading level). Unfortunately, the

DIT can say little about the development of the schema within childhood, except to say that empirically, adolescent and older participants recognized it as, at best, a secondary consideration.

The Maintaining Norms Schema. The Maintaining Norms schema is representative of a society-wide moral perspective. Within the maintaining norms perspective the moral basis of society is understood in terms of how cooperation can be organized on a society-wide basis. However, drawing heavily from the description of Kohlberg's stage 4, the organization of society this schema prioritizes is based on an understanding of rules, roles and the importance of authorities. In addition to Kohlberg's description of stage 4, the Maintain Norms Schema is also informed by conception of the adolescents' developing understanding of political thought and in particular adolescent authoritarianism.

More specifically the Maintaining Norms schema has been defined as having the following characteristics: (a) a perceived need for generally accepted social norms to govern a collective; (b) the necessity that the norms apply society-wide, to all people in a society; (c) the need for the norms to be clear, uniform, and categorical (i.e., that there is "the rule of law."); (d) the norms are seen as establishing a reciprocity (each citizen obeys the law, expecting that others will also obey); and (e) the establishment of hierarchical role structures, of chains of command, of authority and duty (e.g., teacher-pupil, parent-child, general-soldier, doctor-patient, etc).

In short, the Maintaining Norms schema prioritizes the established social order and promotes its maintenance as a moral obligation. Consistent with Kohlberg's stage 4, the Maintaining Norms schema support the view that without law there would be no order, people would act on their own special interests with the result a chaotic and lawless society. This schema, does not provide any additional rationale for defining morality beyond simply asserting that an act is prescribed by the law, is the established way of doing things, or is the established Will of God.

Post-conventional schema. Compared to Kohlberg's view of the postconventional stages, DIT researchers assume a different definition of what constitutes a post-conventional system. Avoiding ties to any given philosophical theory or tradition, DIT researchers describe the essential features of Post-conventional thinking in more general terms. In this view, postconventional thinking suggests all moral obligations are to be based on criteria that

emphasize shared ideals, are fully reciprocal, and are open to scrutiny (i.e., subject to tests of logical consistency, experience of the community, and coherence with accepted practice.

Based on these descriptions, one can observe that the main source of variance in the DIT is provided by the differences between maintaining norms (conventionality) and Postconventionality. These differences are what Kohlberg regarded as the distinction between Stage 4 and Stage 5; and others described as the development of political thought. Although the focus of the DIT measurement system is more directly on the shift from maintaining norms to postconventional thinking than prior models (e.g., Kohlberg's system), the significance of this shift is noteworthy. For instance, the distinction between conventionality and post-conventionality is what tends to drive so many public policy disputes such as the reactions to the wars in Iraq and Afghanistan, how best to stimulate an economy, minority rights, religion in the schools, medical policy, and so on. Further and perhaps most importantly given the events following 9/11, conventional and post-conventional reasoning addresses the divide between religious fundamentalism and secular modernism.

Applying the Neo-Kohlbergian approach to the DIT.

How does the DIT work? With the transition to a moral schema approach we also revisited questions about our measure of moral judgment development, the DIT, and why it works. As many of you know the DIT presents participants with a moral dilemma and then asks them to rate and rank 12 items for each dilemma. Each of the items raise particular issues that define the central features of the dilemma based on different moral schema considerations. These items do not present a complete rationale and interpretation of the dilemma but provide the gist of an explanation using a sentence fragment approach. The sentence fragment approach was adopted because early on in the development of the DIT it was noted that items which contained more detailed interpretations of the dilemmas yielded poor developmental indices in part because these items were prone to reinterpretation and idiosyncratic responding. With the shift to a schema approach it is now more evident why the sentence fragment approach worked better than the other attempts at developing an objective measure of moral judgment development. As we know, schemas are said to capture patterns based on our experiences around particular content areas. It is claimed that we have these moral schemas in order to help us interpret and

understand social situations and are central to how we problem-solve. In short, we see the DIT as an efficient means of triggering moral schemas. That is, sentence fragments are particularly well-suited to trigger a schema because the fragment provides just enough information to suggest an interpretation, and then the individual must fill in the necessary information to fully make sense of the item--the test-taker must meet the item more than half way and fill in the detail. If the item is acceptable to the participant we assume that the item matches the participant's preferred schema and will be rated as important and potentially ranked as most important. However, if the item does not make sense or is viewed as too simplistic, then the item is rated as less important and will not be ranked. In short, DIT researchers assume that the rating and ranking of items across stories provide an index of the participant's preferred schema and more generally, represent how the participant generally approaches moral decisions beyond the DIT; that is taps into the individual's moral schema.

How should the DIT be validated? Because of the transition to a schema approach it became apparent that the strategy for validating the DIT would have to be modified as well. It was no longer possible to refer to Kohlberg's six-stage model or use the validation strategy Kohlberg proposed. As mentioned previously, the Kohlberg group viewed the validity of the standard issue scoring system as the degree to which the data conformed to the theoretical stage model. By giving up this model, a new validation process was required.

The validation process we adopted focused on the two aspects of the Kohlberg model the group considered essential: that the measure describe a phenomenon that is both cognitive and developmental and then expanded these criteria to fit our schema approach. The resulting 6 criteria: (1) differentiation of various age/education groups; (2) longitudinal gains; (3) correlation with cognitive capacity measures; (4) sensitivity to moral education interventions; (5) correlation with behavior and professional decision making; and (6) predicting to political choice and attitude.

Differentiating age/educational groups. The main approach used in these studies is to assess whether or not the DIT is able to distinguish groups which ought to differ on a measure of moral judgment development. For instance, graduate students in political science and philosophy should score higher than other graduate students who are not so well versed in

political and ethical theory. Similarly, college students should score higher than high school students and so on. More recently, large composite samples (thousands of subjects) show that 30% to 50% of the variance of DIT scores is attributable to level of education in samples ranging from junior-high education to Ph.D.s.

Longitudinal gains. The longitudinal gains criteria suggest that a measure of moral judgment development ought to produce evidence of upward movement across time. This criterion follows from the claim that a developmental measure ought to describe change in an upward manner. For instance, a 10-year longitudinal study on the DIT indicates upward change in summary scores for both men and women, for college students and people not attending college, and for people from diverse walks of life. A review of a dozen studies comparing freshman to senior college students (n=755) shows effect sizes (expressed as Cohen's d statistic) of .80 ("large" gains). In short, of all of the variables studied in college student samples, the DIT produces some of the most dramatic longitudinal gains.

Relationship with comprehension measures. Criterion 3 proposes that DIT scores ought to be related to measures of moral comprehension and other cognitive measures. However, relationships with cognitive measures should not be excessive and as such, raise the possibility that DIT scores are actually measuring general cognitive skills. Nor should we find that cognitive measures subsume the relationship between DIT scores and other criterion variables. Overall, the existing literature indicates that DIT scores are significantly related to measures of cognitive capacity and moral comprehension, to recall and reconstruction of post-conventional moral argument, to Kohlberg's measure, and to other cognitive developmental measures.

Sensitivity to moral education interventions. The fourth criterion focuses on whether the DIT is sensitive to specific experiences that ought to stimulate development. Intervention studies are the prototype for this criterion (e.g., presence or absence of a dilemma discussion condition). Findings typically indicate an moderate effect size for dilemma discussion interventions, whereas the effect size for comparison groups was small.

Relationships with prosocial and other outcome variables. The fifth criterion suggests that DIT scores ought to be linked to moral actions and desired professional decision making outcomes.

For instance, one review reports that 32 out of 47 measures of moral action were statistically significant. Furthermore, other reviews have linked DIT scores to many aspects of professional decision-making.

Links with political variables. Criterion six focuses on the link between DIT scores and social/ political variables. In this validity cluster, the assumption is that DIT scores should be significantly linked to political attitudes and political choices. This view follows from the position that the DIT is a measure of macro-morality. As mentioned previously, an understanding of macro-morality addresses an understanding of society-wide institutions and their role in promoting social cooperation through laws and the political process. In a review of several dozen correlates between political attitude and DIT scores it was found that they typically correlate in the moderate range. When DIT scores were combined in multiple regression with measures of cultural ideology, the overall prediction increased to up to two-thirds of the variance in opinions about controversial public policy issues. These issues include abortion, religion in the public school, women's roles, rights of the accused, rights of homosexuals, civil liberties, the rights of minorities, and free speech issues. Given that these issues are among the most hotly debated of our time, the DIT has the potential to contribute to our understanding of individual differences in political preferences and attitudes.

Psychometric support. In addition to these validity criteria, DIT researchers also focused on traditional standards for tests and measures such as acceptable psychometric evidence as well as response stability across different test-taking sets. In addition, DIT scores show discriminate validity from a host of competing variables such as verbal ability/general intelligence and from conservative/liberal political attitudes. Moreover, the DIT is equally valid for males and females since gender accounts for less than one half of a percent of the variance of the DIT.

Locating moral schemas in the Four Component Model.

In the 1990s we also introduced a more nuanced view of the role of moral judgments in moral functioning. As mentioned previously, the Four Component Model proposed that moral action was the product of the four components working together to identify and promote a moral understanding. Located within Component 2, moral schemas were viewed as one of a number of

potential systems an individual might prioritize. Responding to critics who questioned the sufficiency of abstract moral structures in guiding everyday ethical activities, we proposed three levels of moral cognitions. Moral schemas are claimed to be the most general and context-free system for interpreting moral situations. These schemas are labeled as "bedrock schemas" to distinguish the level of assessment provided by the DIT from more context depended interpretive systems. More specifically, the schemas measured by the DIT are viewed as a default system that is evoked when other, more automatic and context-specific, interpretive systems fail or provide incomplete or inconsistent information. By contrast professional codes represent the most concrete level and serve to direct individual behavior in very clearly defined situations. Intermediate concepts fall between these two levels. Unlike codes, the individual must interpret and apply intermediate concepts. Further, intermediate concepts apply to a range or class of situations. However, and unlike the interpretive schema measured by the DIT, intermediate concepts are more narrowly applied and highly contextual.. The interest in operationalizing intermediate concepts is driven by the view that these concepts are more sensitive to educational interventions and more closely related to actions in the targeted context. It is to these measurement systems that we now turn.

The Intermediate concept measurements of moral thinking.

As described above, intermediate concepts are specific concepts that have been identified as central to the ethical life of a professional and have been the subject of significant discussion within the professional field. Typically, these concepts are the primary topics within professional ethics education programs and have significant face validity in the profession. Similarly, at the level of professional ethics education, there is general consensus on what constitutes adequate and inadequate responses to intermediate level concepts. Note that this consensus does not mean that the professionals tend to arrive at a single best application of an intermediate concept in any given situation. Indeed, professionals find it very difficult to arrive at a single best response or definition. However, there is surprisingly good agreement on the *set* of acceptable and unacceptable applications of the concept, e.g., Bebeau & Thoma, (1999). Preliminary work in dentistry supports the utility of intermediate concepts as forming the basis for an assessment of discipline specific moral reasoning. Specifically, Bebeau and Thoma found

support for the sensitivity of the intermediate concept measure to educational interventions in dentistry and for its use as a diagnostic tool.

Although Rest and Narvaez (1994) claim that intermediate concepts are broad-based and provide insight into the moral judgment process, the empirical support for the intermediate concepts measure notion has been limited to young adults in professional programs. This weakness has been noted and some have questioned the claim that ICs define a generalized aspect of the moral reasoning process (e.g., Walker, 2002). To these critics, it is more prudent to view ICs as an artifact of the professional setting and associated well-established set of moral considerations. However, recent research indicates that ICMs may capture moral functioning in the general population.

Characteristics of Intermediate Concept Measures. It is interesting to note how existing intermediate concept measures compare to the traditional moral judgment measures. At first glance there are structural similarities between objective measures of moral judgments and the intermediate concepts measures. Both start with a story to focus the subject's attention, both have different action choices, and provide different justification options to subjects. However upon reflection, the differences between traditional and intermediate concepts measure measures are more striking. First, intermediate concepts measure stories have in common a focus on the target population (e.g., all dilemmas are nested within dentistry, teaching, etc.). Second, multiple possible actions are provided, and in a separate section, multiple justifications. Subjects then rate and rank the appropriateness of items in both sections. Finally and most importantly, intermediate concepts measure s responses are scored in reference to expert opinion (e.g., whether a choice or justification is appropriate). By contrast, moral judgment measures assess item responses by keying each item to a moral stage (e.g., the Defining Issues Test). In the intermediate concepts measure developed by Bebeau and Thoma for dental students, items are ranked as acceptable, neutral and unacceptable based on the majority choices of dentists with ethics training. Given that participants are assessed on choices and justifications, four main scores are generated: the percentage of time subject identified acceptable items as appropriate, and the percentage of time subject select unacceptable items as unacceptable for both action choices and justification items. These scores are then combined to form an overall "good"

(identifying good choices and justifications) and "bad" (identifying bad choices and justifications). Finally, a total score is created combining all four sub areas.

The use of expert choices in place of stage scores is based on the assumption that expert choices represent the application of moral schema to the defining moral issues identified in each story and a sophisticated understanding of current context in which these decisions are made. As such, these choices represent both the expert's "bedrock" ethical concepts, an understanding of the situation, any precedents that may apply, and a general social world-view. Taken together, the expectation is that these highly contextualized decisions are more representative of real-life decision-making and action choices.

Adolescent intermediate concepts measures. Considering character education generally and adolescent populations in particular, it seems reasonable to suggest that an intermediate concept measure may provide better representations of moral thinking within specific contexts particularly salient to adolescents. Furthermore, intermediate concepts measure measures may be more sensitive than traditional measures to the quality of moral thinking and reactions to interventions. Therefore, intermediate concepts measures should represent a particularly informative outcome measure for character education programs interested in assessing moral decision-making. In order to transport the intermediate concepts measure measurement system developed by Bebeau and Thoma to adolescent populations, three main issues need attention: the actual concepts to be studied, the identification of specific dilemmas that capture an intermediate concept, and specific items yoked to each dilemma that represent plausible action choices and justifications. In addition, these items must differ in how appropriatly they reflect the concept.

Identifying the concepts. Regarding the first point, there appears to be a common set of concepts that drive current character education programs. For instance Likona notes two core concepts that should reflect character: respect and responsibility. He further suggests that character incorporates concepts of honesty, fairness, tolerance, prudence, self-discipline, and courage. Comparable lists have been incorporated in other character education programs (e.g., Bonner Foundation Conference papers). Similar to intermediate concepts in the professions (e.g., due process and informed consent), these concepts of character can be viewed as being understood by the individual based on his/her moral judgment processes interacting with

contextual factors including training, experience, precedent, and cultural definitions. Thus, it may be that there is a conceptual overlap between what are called intermediate concepts in the professional literature and aspects of character in the character education literature.

Identifying the stories. Following the decision to frame the measure using the typical content lists of character education programs, the next step in the creation of the AD-icm was to identify appropriate dilemmas that are used to highlight an application of each concept. A number of steps were used to develop these dilemmas and in each step care was taken to solicit input from adolescents in order to maximize the relevance of the resulting topics. First, 50 upper division high school students were asked to review the list of concepts given above and write real-life stories that highlight each concept. The results of this exercise ranged from highly creative and detailed stories to short and stereotypical responses. We then reviewed and sorted the stories by concept looking for common themes and situations. From these sorts, student responses were combined to create a set of stories that were relatively uniform in length and complexity. The resulting stories were then presented to 38 high school seniors and 36 college freshman who were asked to rate each story on realism and plausibility. Further, these students were asked to generate action choices for the story protagonist and supply justifications for these choices. We reviewed the plausibility data and discarded stories that were considered unrealistic. Following this process, we identified seven stories that represent one of the character concepts.

Developing the items. Having identified a set of stories, the next step in the measurement design phase was to develop a list of plausible action choices and justifications for each story. We began with the action choices and justifications identified during the dilemma development phase. These responses were sorted by type and a list of possible items was generated for each story. A small group (n=20) of college freshman reviewed the list of items and rated each proposed action choice and justification on a 5-point scale ranging from highly plausible to highly implausible. In addition, these students were asked to generate choices and justifications that they thought were absent from the lists. We then removed items that had low frequencies, altered some that were problematic and added the student nominations where appropriate. No attempt was made to standardize the number of choices or justifications for each story. Thus, some stories had fewer choices and justifications than others. Guiding this decision was the view

that item realism was more important than methodological elegance and to force an equal number of items increased the risk of including obscure and stilted choices.

Developing the scoring key. Following Bebeau & Thoma, the scoring key was developing using expert decisions about the appropriateness of each action choice and justification. Unlike the professions where expertise can be objectively defined, expertise in adolescent reasoning is more unclear. A number of options for defining experts were considered including teachers, adolescents who have successfully maneuvered through the high school years (e.g., academically and socially), parents, and social scientists who study adolescents. Our eventual choice was graduate students in human development and psychology who had completed an adolescent development course. Given the tendency of parents and teachers to view adolescence and adolescent issues in stereotypical terms, we decided to emphasize social-science expertise over experience and contact. Further, we felt that graduate students were not too removed from the cohort under study and were reasonably expert in their understanding of adolescent development.

As a first step, twenty graduate students were asked to rate each of the AD-icm items as acceptable, unacceptable and neutral. Specifically, these raters were asked to consider the question if the adolescent in the story did what the item suggested would that choice (or justification) be acceptable, unacceptable or neutral. Items with good interrater agreement (75% raw agreement) were assigned the appropriate label. Items falling short of this agreement level, were inspected and reworked as needed. A second sample of 24 students repeated the process. At the end of these two review cycles, all of the action choices and justifications for each story were reliably rated in one of the three categories (i.e., acceptable, unacceptable or neutral).

Developing the AD-icm scoring process. Having developed a set of stories and items, the next step was to construct the measure along with a scoring process. The adopted structure of the measure followed other approaches common to objective measures in moral psychology (e.g., the DIT). Specifically, and after reading the story, we ask participants to rate a set of action choices on a five-point scale (I strongly believe that this is a good choice - I strongly believe that this is a bad choice). After rating each action choice, participants then are asked to rank the three best choices and two worst choices. Following the action choice ranking task, the

participant then rate the justification items on a similar 5-point scale (I strongly believe that this is a good reason - I strongly believe that this is a bad reason). Finally, the justification items are ranked using the same process as before (e.g., identify the 3 best and 2 worst). This process is then repeated for each of the seven stories. Thus, for each story the measure provides the participant's assessment of the best and worst choices and justifications. The primary scoring procedures focus on the ranking data and attend to the appropriateness of the items selected. Generally, higher scores reflect a ranking pattern in which the participants and experts agree. That is, if the participant selects the expert defined acceptable items as the best choices and justifications and in turn, identifies as worst choices and justifications the same way the experts rate the item, then he/she will receive a high score. By contrast, failing to match the experts reduces the scores.

What we have found using the Adolescent ICM.

Age Trends. Empirical work on the Adolescent ICM is relatively recent and focused on exploring how well the measure conforms to theoretical expectations. As with the DIT, there are expectations associated with a measure of moral thinking in the neo-Kohlbergian perspective. Chief among these is the claim that moral thinking ought to be developmental. We now have multiple samples from the US indicate that students do improve in their ability to identify better and worse choices and justifications. These studies are corroborated by a European sample, which, when taken together, indicates that the measure is sensitive to age educational groups across the high-school years.

Differences between better and worse scores. Within these general trends we find that across measures and samples representing different age groups, identifying the "bad" choices/justifications lags behind the "good" items. Although speculative, the difficulty associated with identifying bad choices and justifications may be simply a reflection of socialization and training where the emphasis is on the acceptable and good. Thus, students may be more on their own when it comes to deducing bad choices and justification. Whether this finding is a reflection of how our culture socializes its children or due to a more general developmental process, a continued focus on the difference between identifying good and bad choices seems especially warranted.

Gender differences. Further inspection of these general age and education trends on the Adolescent ICM indicated a surprisingly large gender difference favoring women. Years of work on the DIT have found very small gender effects (also favoring women), and thus we assumed that gender would not be a large contributor to difference on the ICM measure. It may be that large gender differences are due to the fact that woman have an advantage because the stories developed for the AD-icm were more influenced by their input. Although we have no data to support this claim, our recollection is that women responded to our early requests for stories and items with more detailed responses. To counter this possibility we were sensitive to gender in the reactions to the various stories during the development phase of the measure, however it still may be that the dilemmas eventually selected are more familiar to women. Future research using the adolescent ICM should monitor and attend to the possibility of gender differences.

Relationships with moral action. Central to the neo-Kohlbergian model is the view that measures of moral thinking ought to be related to action. Indeed, the development of the Four Component Model was an attempt to spell out how moral thinking relates to moral action. Similar expectations apply to the adolescent ICM. As a preliminary step in assessing the relationship between ICM scores and action we focused on inappropriate behaviors within the school setting. Specifically we identified students who had been placed in in-school suspension. Typically, to be placed in suspension students have a history of making poor choices as they have exhausted all of their first and second chances at remediation. Given this pattern of bad choices we expected these students to have a similar difficulty in identifying acceptable and unacceptable items on the ICM measure. Our findings support this notion. Although the group of students under suspension included a range of educational levels, ICM scores placed them significantly below our youngest group. Thus we have evidence that students who objectively make bad choices also underperform on the ICM measure. Interestingly, a recent European study finds a similar difference between typical and acting out students on a translated version of the ICM. Taken together, these findings support theoretical expectations that ICM measures ought to relate to behavior.

Relationships with the DIT. Validating an ICM measure of moral thinking includes support for the claim that the measure relates to other measures within the moral domain. The preliminary

evidence suggests that the Adolescent ICM is related to the DIT albeit at a moderate level. However, the overall relationship masks some interesting patterns. Particularly noteworthy is the negative association between the personal interest schema and ICM scores. Recall that the personal interest schema emphasizes the self and personal relationships in reasoning about moral situations and does not attend to more system-wide considerations of cooperation. We see these findings as supporting the view that a personal interest orientation is a liability for understanding intermediate concepts in a way that is consistent with our expert key and the norms they represent. That is, only when social norms, laws, and principles are featured in one's moral thinking does the application of intermediate concepts approach the prevailing view represented by our expert key. We note that these findings further support the utility of both the DIT and ICM measurement systems and clarify the role each measurement system plays in Component II of the Four Component Model.

Current status and interests.

The neo-Kolbergian model is now over 13 years old and it is possible to summarize its impact on the field. Clearly the DIT is still a force in the profession with over 30 thousand participants using the measure each year. As such, the DIT is the measure of choice in intervention studies, evaluations of ethics programs and college student outcome assessments. The DIT is also used to support the construct validity of newer more specialized measures. Furthermore and given that the measurement has been stable for so long, the DIT is also used to track trends in moral judgment development in the general population both in the US and elsewhere.

Additionally, there has been a consistent focus on the development of intermediate concept measures. Intermediate concept measures are typically well-received and have high credibility with both participants and consumers of the resulting data. At present there are approximately a dozen measures in various stages of development and we see this as an area of particular growth. The adolescent ICM is perhaps the most broadly applicable of these various measures.

More generally, and consistent with the traditional focus of Rest and his colleagues, the Neo-Kohlbergian model has promoted the development of many measures that capture various aspects of moral functioning including measures of moral sensitivity, judgment, and motivation. We fully expect that these trends will continue in the near future.

For further information on the neo-Kohlbergian approach:

Bebeau, M. & Thoma, S. (1999). Intermediate concepts and the connection to moral education. Educational Psychology Review, 11, 343-360.

Rest, J.R., Narvaez, D., & Bebeau, M. & Thoma, S.J. (1999). *Postconventional Moral thinking: A neo-Kohlbergian approach*. L. Earlbaum: Mawah, NJ

Rest, J.R., Narvaez, D., Thoma, S.J & Bebeau, M. (2000). A neo-Kohlbergian approach to moral judgment: Overview of the Defining Issues Test Research. *Journal of Moral Education*, 29, 381-397.

Thoma, S.J. (2006). Research using the Defining Issues Test. In Killen and Smetana (Eds.), *Handbook of Moral Psychology*. L. Earlbaum: Mawah, NJ.

Thoma, S.J., Derryberry, P., & Crowson, H. M. (2013). Describing and testing an intermediate concept measure of adolescent moral thinking. *European Journal of Developmental Psychology (special issue). Vol*. 10, 239-252.