Article

T·R·E

Theory and Research in Education
2014, Vol. 12(3) 347–365
© The Author(s) 2014
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1477878514545208
tre.sagepub.com
SAGE

Measuring moral thinking from a neo-Kohlbergian perspective

Stephen J. Thoma
The University of Alabama, USA

Abstract
The neo-Kohlbergian model revises and extends Lawrence Kohlberg's model of moral reasoning development to better reflect advances in research and theory. In moving from Kohlberg's global stage model to a multi-process description of moral functioning, these modifications are most evident in the ways in which moral thinking is described, measured, and interpreted. This article highlights the two primary measurement systems of moral thinking associated with the neo-Kohlbergian perspective: measures of moral judgment development and the more recently identified intermediate concepts measures. In describing both systems, attention is given to the nature of the information supplied and evidence used to support them.

Keywords
Measurement systems, moral thinking, neo-Kohlbergian model, virtues

History and background

How individuals reason about moral issues represents the central question driving research in the cognitive developmental tradition. Beginning with Piaget (1932/1965) and later Kohlberg (1969), this question has been addressed by attending to the individual's developing understanding of cooperation and associated judgments of fairness. Kohlberg and his colleagues believed that the development of moral judgments was best described by broadly defined stages that encompassed a person's sensitivity to moral issues, moral emotions, as well as the motivation to act. In the 1980s, the support for a global stage view began to erode both within the field of moral psychology and developmental psychology more generally (Turiel, 2006). In place of the global stage view,

Corresponding author:
Stephen J. Thoma, Educational Psychology, The University of Alabama, 310 Carmichael Hall, Box 870231, Tuscaloosa, AL 35487, USA.
Email: sthoma@bamaed.ua.edu

moral psychology turned to models of moral functioning that treated these various aspects of development as separate processes. In addition to unpacking moral functioning, the process of turning away from global stages has had the effect of broadening psychological understanding of the moral domain into areas beyond judgments of fairness. For the neo-Kohlbergian model, this expansion took the form of the four component model (hereafter the FCM; Rest, 1983).

According to the FCM, moral actions are the result of a least four component processes operating individually and in interaction. The 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. These component processes, although linked, are assumed to be conceptually distinct and may develop at different rates. 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 is 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.

To highlight the shift away from Kohlberg's global stages, the FCM assumes that the moral judgment construct is located within Component 2 and, in contrast to Kohlberg's view, does not provide direct information on affiliated constructs such as moral sensitivity (Component 1) or motivation (Component 3). Furthermore, even within Component 2, it is only one of many strategies an individual can use to construct an idealized response to a moral event. For instance, Rest noted that in addition to moral judgment processes, one might prioritize social norms, religious prescription, or something else. Thus, moral judgment processes in the FCM hold an important, yet much more modest, position within moral functioning than they did in Kohlberg's model.

The adoption of the FCM set the stage for the more significant modifications that were proposed in the late 1990s–2000s. Chief among these changes was the adoption of a schema view of moral judgment development. The use of schemas to define a model of moral judgment development signaled significant differences in the focus of the model and assessment process. To contrast Kohlberg's view with the evolving neo-Kohlbergian perspective, Kohlberg's moral stages were described in terms of cognitive operations that directly describe the structure of moral thinking. The measurement system associated with Kohlberg's method was designed to hone in on the structural features of one's thinking while avoiding any interference from the specific features, or content, of the target situation. That is, when discussing a moral dilemma, the situation, protagonist roles, and so on are viewed as content and are not central to the assessment process. Using this approach, the researcher assumes that he or 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, the neo-Kohlbergian view suggests 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 this view, a more appropriate conception of what develops in moral thinking is represented by a schema view

that highlights how individuals understand, organize, and prioritize moral content such as societal norms, systems, and organizations.

In addition to revising the moral judgment construct as a developmentally ordered sequence of moral schemas, attention was also given to where these schemas fit within the FCM's conception of Component 2. As mentioned previously, Component 2 is claimed to encompass multiple interpretive systems by which the individual evaluates and judges socio-moral situations. The question addressed by the neo-Kohlbergians was whether these interpretive systems also differed in their coverage from the most general and abstract to systems that were concrete and tied to particular contexts. In part, attention to questions of broad versus narrow systems was prompted by critics of the cognitive developmental tradition who questioned the sufficiency of abstract moral structures in guiding everyday ethical decisions and as a basis for evaluating ethics curricula in professional schools (e.g. Strike, 1982). What was needed, they argued, were theoretically grounded assessments more directly tied to the professional context.

To clarify the organization of Component 2, proponents of the FCM recast moral schemas as the most general and context-free system for interpreting moral situations. These schemas were labeled as 'bedrock schemas' to distinguish the level of assessment provided by the moral judgment measures from the more context-dependent interpretive systems. More specifically, a moral schema is viewed as a default system that is evoked when other, more automatic and context-specific, interpretive systems fail or provide incomplete or inconsistent information (Rest et al., 1999; Thoma, 2006). In contrast to these bedrock schemas, neo-Kohlbergians identified professional codes as the most concrete level. As described within the revised description of Component 2, codes of conduct are claimed to direct individual behavior in very clearly defined situations and with minimal need for interpretation.

Between the bedrock schema and codes of conduct, Rest and his colleagues identified a new level of abstraction labeled intermediate concepts. Judgments at the intermediate concept level are understood to require more interpretation in order to apply them than would be required in order to apply a professional code of ethics. One cannot simply follow a given script. Furthermore, intermediate concepts apply to a range or class of situations and are not tied to a particular triggering event or situation. For example, in the professions, intermediate concepts have included informed consent, patient confidentiality, and beneficence (e.g. Bebeau and Thoma, 1999). It is interesting to note that intermediate concepts are the primary foci of professional ethics education and typically represent ethical considerations that are of central concern to the field. Contrasting intermediate concepts to moral schemas as assessed by measures of moral judgment development, intermediate concepts are more narrowly applied and highly contextual. A primary 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.

In summary and using its current formulation, the FCM points to two levels of moral thinking that have broad-based utility: 'bedrock schemes', which highlight the basic strategies by which the individual understands cooperation and fairness, and intermediate concepts, which focus on how the individual interprets and reacts to moral situations in context. Both of these levels of moral thinking have been operationalized and have

established measurement systems associated with them. The remaining sections of this article provide a description of each system and the resulting picture of moral thinking each presents.

Measures of moral schema

The measurement of moral schemas in the neo-Kohlbergian model is conceptually the most directly linked to Kohlberg's original view of moral judgment development. It is primarily this link that prompted the neo-Kohlbergian label used to represent the research tradition associated with Rest and his colleagues. Specifically, neo-Kohlbergians have relied on the Defining Issues Test (hereafter the DIT; Rest, 1979) to measure moral schemas. The DIT is an objective paper-and-pencil measure of moral judgment development created by James Rest who had a long association with the Harvard group and Kohlberg in particular. In developing the DIT, Rest used moral dilemmas that originated from Kohlberg's work. Similarly, the specific items used on the DIT were distillations of participant interview responses on the Kohlberg interview. Indeed, the DIT gained early acceptance primarily because of its close relationship to the Kohlberg approach. At first, the view in the field was that Rest and his colleagues had developed a 'quick and dirty' objective measure of Kohlberg's theory which might be helpful when one could not use the preferred Kohlberg interview approach. Although the ties between the two traditions are now primarily historical, there are some basic overlapping assumptions that highlight the need to maintain the connection.

Similarities with Kohlberg's theory

The basic framework of the neo-Kohlbergian model significantly borrows from Kohlberg's approach. First, central to both perspectives is a focus on cognition. Like Kohlberg, the neo-Kohlbergians 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. Second, and consistent with the Piagetian perspective, the two traditions agree that the individual does not passively accumulate information about the social world. Instead, social information is self-constructed and organized. Third, both traditions 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, both traditions believe that 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 that to be moral, conventions must conform to a shared ideal 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. To neo-Kohlbergians, 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 describes the individual's understanding of the moral basis of laws, governing structures, and general practices of society. In the macro-morality perspective, the idealized perspective is one that prioritizes impartial principle over partisanship and 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 describes how the individual emphasizes positive interpersonal characteristics in interactions with special individuals. Clearly, there are both tensions and communalities between these conceptions. The neo-Kohlbergians have argued, however, that Kohlberg's theory is a better description of macro-moral thinking and they focus on macro-moral thinking in their measurements of moral schemas.

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. Many of these differences were presented in the preceding sections describing the shift away from Kohlberg's global stage model to the current FCM. However, at the measurement level, an additional difference is evident in the priority placed on verbal data. It is common to read how verbal data are 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. Noting more recent trends in cognitive science, neo-Kohlbergians conclude that there is little support for the view that participants have any insight into the processes that produce their judgments. Indeed, there seems little in this literature 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 other measurement systems including the DIT. To explain this difference, it may be that individuals are able to recognize the superiority of post-conventional reasoning without being able to articulate and defend them. That is, individuals may rely on tacit knowledge to determine a preferred strategy. In the neo-Kohlbergian view, there is a utility in focusing on tacit knowledge as it may be the more influential system used to reach decisions and justify moral actions within real-life contexts.

In addition to differing views on the importance of verbal data, the neo-Kohlbergians also deviate from Kohlberg on the claim of a universal moral system. 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 assumed to be universal. Additionally, Kohlberg argued that a universality claim was essential to avoid moral relativism, which might allow for communities to define a moral system in any way they wanted (e.g. his often-quoted question used to

highlight this concern – 'so cannibalism is fine for cannibals?'). Our shift toward a schema 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 offered by neo-Kohlbergians reduces the universality claim to an empirical question. It is presupposed 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 arrangements, and current concerns are debated within the community and become shared experiences which inform individual moral thinking. In describing this view, neo-Kohlbergians draw parallels between moral understanding and common law, which not only shares some common principles across cultures but also some unique features based on the specific experiences of the various communities. It is further assumed that common morality changes over time and as with other systems such as law and science, common morality evolves as new precedents and data are assimilated.

How is development defined in the Neo-Kohlbergian model?

In addition to altering the developmental model underlying the measure, the neo-Kohlbergians 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 1970s. More recently, however, the fit of Kohlberg's model to DIT data has been reassessed. 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 has 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. Taken together, the best fitting structure using DIT data is no longer the six Kohlberg stages. Instead, a three-cluster model loosely informed by Kohlberg's system seems more appropriate.

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 schemas are labeled as follows: the Personal Interest 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 ninth-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 Maintaining Norms schema is also informed by a 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 supports the view that without law there would be no order, people would act on their own special interests, with the result that a chaotic and lawless society would ensue. 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 post-conventional 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 their view, post-conventional thinking emphasizes the position that moral obligations are to be based on criteria that prioritize 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

post-conventionality. 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 post-conventional 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 consistently related to political choices and voting behavior (Thoma, 2006) and has been shown to distinguish conservative and liberal religious perspectives (e.g. Narvaez et al., 1999).

Applying the Neo-Kohlbergian approach to the DIT

How does the DIT work? With the transition to a moral schema approach, neo-Kohlbergians also revisited questions about the DIT and why it works. Briefly, 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 poorly performing 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. Moral schemas are claimed to exist 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. The individual's role is to then fill in the necessary information to fully make sense of the item – the test-taker must meet the item more than half way. 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.

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 adopted validation process 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 expand these criteria to fit a schema approach. The resulting six criteria are as follows: (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 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%–50% of the variance of DIT scores is attributable to level of education in samples ranging from junior-high education to PhDs.

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 (e.g. Pascarella and Terenzini, 2005).

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 one 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 to reconstruction of post-conventional moral argument, to Kohlberg's measure, and to other cognitive developmental measures (e.g. Thoma et al., 1999b).

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 a moderate effect size for dilemma discussion interventions, whereas the effect sizes for comparison groups are small (e.g. Thoma, 2006).

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 related. Furthermore, other reviews have linked DIT scores to many aspects of professional decision making (e.g. Rest, 1986; Thoma, 2006).

Links with political variables. Criterion 6 focuses on the assumed link between DIT scores and social/political variables. The claim that a relationship should exist follows from the position that the DIT is a measure of macro-morality. As mentioned previously, macro-morality addresses the individual's 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.

It has been claimed that the link between the DIT and socio-political variables is simply the result of a liberal political bias in the measurement system (e.g. Emler et al., 1983). In short, these critics argue that conventional and post-conventional moral strategies represent conservative and liberal political worldviews found in the general adult population. They hold that by applying his theoretical (and personal) biases, Kohlberg placed these two strategies in a developmental order prioritizing the liberal view. These critics go on to suggest that when measures like the DIT report correlations with political variables, the results are simply due to the political orientation of the respondent and not to the individual's moral judgment strategies. A number of articles have tested these claims empirically and generally find that political orientation cannot account for the relationships between DIT scores and socio-political judgments (Thoma et al., 1999a, 1999b).

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 (Thoma, 1986).

Current status of the measure

After 40 years, the DIT remains a force in the profession with over 30,000 participants using the measure each year (Center for the Study of Ethical Development Reports,

2014). The majority of these individuals are participants in basic social science research projects, evaluations of ethics programs, and college student outcome assessments. More recently and because the measurement system has been stable for so long, the DIT has been used to track general population trends in moral judgment development within the United States and elsewhere (Thoma et al., 2014). In addition to these traditional uses, the DIT is also employed to support the construct validity of newer more specialized measures. Of these measures supported by the DIT, one set comprises the intermediate concept measures (ICMs). In the following section, this second major class of moral thinking measures is described with a particular focus on the adolescent ICM.

The intermediate concept measurements of moral thinking

As described in the opening sections of this article, measures based on intermediate moral concepts are described as assessments of moral thinking that are nested within a particular social context. Although Rest and Narvaez (1994) claim that intermediate concepts are broad-based and provide insight into normative moral thinking, until recently, the empirical support for the ICM approach has been limited to young adults in professional programs (e.g. Bebeau and Thoma, 1999). This weakness has been noted and some have questioned the claim that intermediate concepts define a generalized aspect of the moral reasoning process (e.g. Walker, 2002). To these critics, it is more prudent to view intermediate concepts as an artifact of a clearly defined professional context and associated well-established set of moral considerations. More recently, Thoma et al. (2013) have established an ICM that applies to adolescent moral thinking in the general population. Given that ICMs are relatively new measurement systems, the development of the adolescent ICM is presented below in some detail along with the preliminary empirical evidence supporting its interpretation.

Characteristics of ICMs. It is interesting to note how existing ICMs compare to the traditional moral judgment measures. At first glance, there are structural similarities between objective measures of moral judgments and the ICMs. Both start with a story to focus the subject's attention, and both provide different action choices and justifications options to subjects. However, upon closer examination, the differences between traditional measures and ICMs are more striking. First, ICM stories have in common a focus on the target population (e.g. all dilemmas are nested within the adolescence experience, 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, ICM 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 schema (e.g. the DIT). In the prototype ICM developed by Bebeau and Thoma (1999) for dental students (and later translated into the adolescent context by Thoma et al., 2013), 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 a subject identified acceptable items as appropriate, and the percentage of time a subject selected unacceptable items

as inappropriate, for both action choices and justification items. These scores are then combined to form overall acceptable (identifying acceptable choices and justifications) and unacceptable scores (identifying unacceptable choices and justifications). Finally, a total score is created combining all four sub-areas.

The use of expert choices in place of theoretically defined scores is based on the assumption that expert choices represent the application of moral schemas to the defining moral issues identified in each story coupled with a sophisticated understanding of current context in which these decisions are made. As such, these choices represent the expert's 'bedrock' ethical concepts, an understanding of the situation, any precedents that may apply, and a general social worldview.

Adolescent ICMs. Considering adolescent populations in particular, it seems reasonable to suggest that an ICM may provide better representations of moral thinking within specific contexts particularly salient to adolescents. Furthermore, ICMs may be more sensitive than traditional measures to interventions designed to influence adolescent moral thinking such as character education programs. In order to translate the intermediate concepts measurement system developed by Bebeau and Thoma (1999) to adolescent populations, three main issues needed 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, items must differ in how appropriately they reflect the concept.

Identifying the concepts. Typically, intermediate concepts are identified by noting the ethical issues discussed by professionals within the governing bodies. They tend to be the primary topics within professional ethics education programs and have significant face validity in the profession. In the absence of a well-described set of ethical considerations, Thoma et al. (2013) identified adolescent intermediate concepts by turning to character educational programs. These programs are typically designed to influence the ethical thinking and broader ethical formation of students in middle and high school. Across these educational programs, the majority use the virtues to guide instruction. In describing the focus of character education programs, Lickona (1991) notes two core concepts that should reflect character: the virtues of respect and responsibility. He further suggests that good character incorporates concepts of honesty, fairness, tolerance, prudence, self-discipline, and courage. Comparable lists have been incorporated in other character education programs (e.g. Arthur, 2008). Similar to intermediate concepts in the professions (e.g. due process and informed consent), these concepts of character can be viewed as requiring interpretation based on one's moral judgment and contextual factors. Overall, Thoma et al. (2013) suggest 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 Adolescent ICM (AD-icm) was to identify appropriate dilemmas 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. These stories were then reviewed and sorted by concept with attention to 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 freshmen who were asked to rate each story on realism and plausibility. Furthermore, these students were asked to generate action choices for each story's protagonist and supply justifications for these choices. The plausibility ratings were reviewed and unrealistic stories were discarded. Following this process, seven stories were identified for each character concept.

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. Action choices and justifications identified during the dilemma development phase became the starting point for item construction. These responses were sorted by type and a list of possible items was generated for each story. A small group (n = 20) of college freshmen reviewed the list of items and rated each proposed action choice and justification 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. From these responses, items were either removed or altered. Additional items were considered based on the student nominations. 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 the decision to allow item numbers to vary was the position that item realism was more important than simple methodological consideration coupled with a concern that to force an equal number of items increased the risk of including obscure and stilted choices.

Developing the scoring key. Following Bebeau and Thoma (1999), the scoring key was developed 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 ambiguous. 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, it was decided to emphasize social science expertise (Eccles et al., 1996). Additionally, it was noted that graduate students were not too removed from the cohort under study and were reasonably expert in their understanding of the adolescent context.

As a first step, 20 graduate students were asked to rate each of the AD-icm items as acceptable, unacceptable, and neutral. Specifically, these raters were asked to consider

whether it would be acceptable, unacceptable, or neutral if a hypothetical adolescent selected that choice (or justification). Items with good inter-rater agreement (80% 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, participants are asked to rate a set of action choices. 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 rates and ranks the justification items in a similar manner. 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 or she will receive a high score. By contrast, failure to match the experts' ratings reduces the scores.

What has been 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. There are now multiple samples from the United States and Europe, which indicate that students do improve in their ability to identify better and worse choices and justifications (Frichand, 2011; Thoma et al., 2013). Together, these findings indicate that the measure is sensitive to age educational groups across the high school years.

Differences between acceptable and unacceptable scores. Within these general trends, we find that across measures and samples representing different age groups, identifying the 'bad' choices and justifications lags behind decisions about the 'good' items. Although speculative, the difficulty associated with identifying bad choices and justifications may be 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 is 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 large gender difference favoring women. This finding is a surprise based on assumptions formed from the extensive literature on gender differences on the DIT where only small gender effects are noted (also favoring women; e.g. Thoma, 1986). The straightforward interpretation of this finding is that the observed gender differences indicate a developmental advantage for women across the adolescent years. To complement this interpretation, it is interesting to note that in recent years the gender difference on the DIT is growing in late adolescent populations (Center for the Study of Ethical Development Reports, 2014). However, another possibility for the disparity between the adolescent ICM and DIT findings is methodological. It may be that women have an advantage because the stories developed for the AD-icm were more influenced by their input. Although speculative, Thoma et al. (2013) note that women were more responsive to requests for stories and items, and provided more detailed responses during the measurement construction phases. To assess the possibility that the measure was more friendly to women, gender differences in the reactions to the various stories and items were monitored; however, it still may be that the dilemmas eventually selected advantage women.

Relationships with moral action. Central to the neo-Kohlbergian model is the view that measures of moral thinking ought to be related to behavior. Indeed, the development of the FCM was an attempt to describe how moral thinking relates to moral action. Similar expectations also apply to the adolescent ICM. The current evidence supporting a relationship between ICM scores and action focuses on inappropriate behaviors within the school setting. Specifically, students who had been placed in in-school suspension were compared to their peers without this record of acting out. Typically, to be placed in suspension, students must have a history of making poor choices as they have exhausted all of their first and second chances at remediation. Given a pattern of bad choices, it was expected that these students would be more likely to have a similar difficulty in identifying acceptable and unacceptable items on the ICM. Across samples, the findings support this notion. Although the group of students under suspension included a range of educational levels, ICM scores placed them significantly below the youngest group. Thus, there is evidence to suggest that students who objectively make bad choices also underperform on the ICM. Interestingly, a recent European study found a similar difference between typical and acting out students on a translated version of the ICM (Frichand, 2011). 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 evidence that it is related to other measures within the moral domain. Using the DIT as the established measure within the moral domain, obtained relationships with the AD-icm indicate a moderate association. However, the overall relationship masks some interesting patterns. Particularly noteworthy is the negative association between the Personal Interest schema and ICM scores. Indeed much lower AD-icm scores were associated

with individuals who prioritized the Personal Interest schema (Thoma et al., 2013). As described earlier, the Personal Interest schema emphasizes interests of the self and the importance of personal relationships in reasoning about moral situations. Lacking from this perspective is the attention to more system-wide considerations of cooperation. These findings support the view that a personal interest orientation is a liability in understanding intermediate concepts as defined by our expert key and the norms they represent. Only when social norms, laws, and principles are prioritized in one's moral thinking does the application of intermediate concepts approach the prevailing view represented by our expert key.

Summary

The overall picture of adolescent moral thinking presented by the AD-icm suggests that there is a growing ability to apply the virtues across the adolescent years. Given the findings using the more traditional measures of moral judgment development (e.g. King and Mayhew, 2002; Rest, 1986), it is likely that this growth is tied to normative experiences associated with the school or social context, experiences which may be enhanced by practitioners. Additionally, it may be that development in adolescents' general moral schema may be driving growth on the AD-icm through a shift from a personal interest to maintaining norms perspective. The latter finding highlights the relationships between the different levels of moral thinking presupposed by the FCM and provides some suggestions about the kinds of information and processes that inform the application of virtue concepts.

Of additional interest is the difference in adolescents' ability to identify acceptable and unacceptable choices and justifications. This finding – across samples and at least two cultures – implies that during the adolescent years, these two types of choices are not simply mirror images of the same concept and ought to be treated independently. A difference in ability to identify positive and negative applications of these virtues concepts could be linked to the ways in which they are presented to children. Consistent with this view, the AD-icm findings reflect the emphasis on positive examples of a virtue in both the formal and informal teaching process. At the very least, these data suggest that when designing educational interventions, practitioners ought to pay increased attention to inappropriate applications of virtue concepts and in so doing help the adolescent come to understand both dimensions of the concept.

A focus on both acceptable and unacceptable applications may be particularly important for younger adolescents because of the finding that this gap is largest in the acting out samples and younger adolescents and then tends to decline – but never disappear – in older groups. Thus, the effect is likely to be developmental. If so, one could argue that the coordination of acceptable and unacceptable choices is an ongoing feature of adolescent thinking and a marker for a growing maturity in moral thinking.

Finally, the large gender difference associated with the AD-icm stands out in the moral domain and may indicate a particular advantage for women in reasoning about socio-moral issues within context. Although gender differences are small on measures of broad-based moral thinking, perhaps the move to contextual assessments magnifies the

advantage by tapping into other strengths on associated constructs such as moral sensitivity (Bebeau et al., 1985). Conversely, the observed gender differences may represent different developmental trajectories rather than different attributes. Regardless of the cause, gender differences in the application of virtue concepts should be expected within age-based group settings.

Current status of the ICM approach

The development of the adolescent ICM is part of a growing interest in these measures. Taken together, ICMs 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 this appears to be an area of particular growth. In addition to its use as an outcome measure, Roche et al. (2014) show that ICMs can also form the basis of an intervention. In her study, Roche et al. (2014) introduced pharmacists to an ICM and had groups react and defend various choices and justifications. The outcome assessments of the intervention supplied by participants highlighted the interest practitioners have in the ICM approach and its utility as the focal point of a professional ethics intervention.

Conclusion

In the neo-Kohlbergian model, it is claimed that measures of moral thinking differ in their level of abstraction. Measures of moral schema represent the most abstract cognitions and serve to define the default systems used to interpret and justify a moral perspective. Although there is great utility in knowing how an individual reasons about cooperation and fairness, the model makes clear that this level of assessment is not sufficient to capture all of the ways in which moral thinking informs moral decisions. To fill the gap between measures of moral schemas and more contextual assessments, researchers from this tradition have identified intermediate concepts to capture moral thinking in context. This identification of intermediate concepts does not invalidate or render obsolete measurement systems like the DIT. Instead, the ICMs should be viewed as complementing the traditional systems. Indeed, current evidence supports the view that both systems provide important information and it suggests some interesting relationships between them. Although practitioners may find the ICMs most appropriate for program assessments, any evaluation of a population or setting can benefit from information supplied by both systems.

More generally, the two approaches to the assessment of moral thinking described in this article highlight the traditional focus of researchers associated with the neo-Kohlbergian perspective. From the creation of the DIT, through the development of the FCM to the current neo-Kohlbergian model, the primary objective of this group is to describe features of moral functioning with particular attention to measurement development. In so doing, the Neo-Kohlbergian model has promoted the development of many assessments that capture various aspects of moral functioning, including measures of moral sensitivity, judgment, and motivation. These measures have enriched the field and there is every indication that these trends will continue in the near future.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

Arthur J (2008) Traditional approaches to character education in Britain and America. In: LP Nucci and D Narvaez (eds) *Handbook of Moral and Character Education*, pp. 80–98. New York: Routledge.

Bebeau M and Thoma S (1999) Intermediate concepts and the connection to moral education. *Educational Psychology Review* 11: 343–360.

Bebeau MJ, Rest JR and Yamoor CM (1985) Measuring dental students' ethical sensitivity. *Journal of Dental Education* 49(4): 225–235.

Center for the Study of Ethical Development Reports (2014) http://ethicaldevelopment.ua.edu

Eccles JS, Lord S and Buchanan CM (1996) School transitions in early adolescence: What are we doing to our young people? In: JA Graber, J Brooks-Gunn and A Peterson (eds) *Transitions through Adolescence*, pp. 251-284. Mahwah, NJ: Lawrence Erlbaum Associates.

Emler N, Renwich S and Malone B (1983) The relationship between moral reasoning and political orientation. *Journal of Personality and Social Psychology* 45: 1073–1080.

Frichand A (2011) Examining moral thinking of adolescents through intermediate concepts. *Psiholoska Istrazivanja* 14: 67–84.

King PM and Mayhew MJ (2002) Moral judgment development in higher education: Insights from the Defining Issues Test. *Journal of Moral Education* 31: 247–270.

Kohlberg L (1969) Stage and sequence. The cognitive- developmental approach to socialization. In: DA Goslin (ed.) *Handbook of Socialization Theory and Research*, pp. 347–380. Chicago, IL: Rand McNally.

Lickona T (1991) *Educating for Character*. New York: Bantam.

Narvaez D, Getz I, Thoma SJ, et al. (1999) Orthodoxy, progressivism and development of moral judgment. *Developmental Psychology* 35: 478–488.

Pascarella ET and Terenzini PT (2005) *How College Affects Students A Third Decade of Research*. San Francisco, CA: Jossey-Bass.

Piaget J (1932/1965) *The Moral Judgment of the Child* (trans. M Gabain). New York: Free Press.

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

Rest J (1983) Morality. In: PH Mussen (ed.) *Handbook of Child Psychology: Cognitive Development*, vol. 3 (edited by J Flavell and E Markman), 4th edn, pp. 556–629. New York: Wiley.

Rest JR (1986) *Moral Development: Advances in Research and Theory*. Minneapolis, MN: University of Minnesota Press.

Rest JR and Narvaez D (1994) *Moral Development in the professions: Psychology and Applied Ethics*. Mahwah, NJ: Lawrence Erlbaum Associates.

Rest JR, Narvaez D, Bebeau M, et al. (1999) *Postconventional Moral Thinking: A Neo-Kohlbergian Approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Roche C, Thoma SJ and Wingfield J (2014) From workshop to e-learning: Using technology-enhanced 'Intermediate Concept Measure' as a framework for pharmacy ethics education and assessment. *Pharmacy* 2014(2): 137–160. DOI:10.3390/pharmacy2020137.

Strike KA (1982) *Educational Policy and the Just Society*. Chicago, IL: University of Illinois Press.

Thoma SJ (1986) Estimating gender differences in the comprehension and preference of moral issues. *Developmental Review* 6: 165–180.

Thoma SJ (2006) Research using the Defining Issues Test. In: M Killen and JG Smetana (eds) *Handbook of Moral Psychology*, pp. 67-92. Mahwah, NJ: Lawrence Erlbaum Associates.

Thoma SJ, Barnett R, Rest JR, et al. (1999a) Political identity and moral judgment development using the Defining Issues Test. *British Journal of Social Psychology* 38: 103–111.

Thoma SJ, Bebeau M and Dong Y (in preparation) Is moral judgment competency declining over time? Evidence from thirty years of Defining Issues Test data.

Thoma SJ, Derryberry P and Crowson HM (2013) Describing and testing an intermediate concept measure of adolescent moral thinking (special issue). *European Journal of Developmental Psychology* 10: 239–252.

Thoma SJ, Narvaez D, Rest J, et al. (1999b) Does moral judgment development reduce to political attitudes or verbal ability: Evidence using the Defining Issues Test? *Review of Educational Psychology* 11: 325–342.

Turiel E (2006) The development of morality (revised edition). In: W Damon and RM Lerner (eds) *Handbook of Child Psychology: Social, Emotional, and Personality Development*, vol. 3 (edited by N Eisenberg), 6th edn, pp. 863- 932New York: Wiley.

Walker L (2002) The model and the measure: An appraisal of the Minnesota approach to moral development. *Journal of Moral Education* 31: 353–367.

Author biography

Stephen J. Thoma is a Professor of Educational Psychology and Director of the Center for The Study of Ethical Development at the University of Alabama. He has written extensively on issues in the psychology of morality with a particular focus on neo-Kohlbergian models of moral functioning. He has been President of the Association for Moral Education and is a Fellow of the American Educational Research Association. He was named the Paul W. Bryant Research professor in 2010 and received the career award from the Association for Moral Education in 2004.