

# How not to evaluate a psychological measure: Rebuttal to criticism of the Defining Issues Test of moral judgment development by Curzer and colleagues

**Stephen J Thoma**

The University of Alabama, USA

**Muriel J Bebeau**

University of Minnesota, USA

**Darcia Narvaez**

University of Notre Dame, USA

## Abstract

In a 2014 paper in *Theory and Research in Education*, Howard Curzer and colleagues critique the Defining Issues Test of moral judgment development according to eight criteria that are described as difficulties any measure of educational outcomes must address. This article highlights how Curzer et al. do not consult existing empirical evidence, misunderstand the model and method associated with the Defining Issues Test, and thereby reach conclusions that are unwarranted, incomplete, and ultimately indefensible. To address these shortcomings, we present an overview of the Defining Issues Test and note relevant criteria for evaluating a measure of moral judgment development.

## Keywords

Moral judgment development, neo-Kohlbergian model, the Defining Issues Test

---

## Corresponding author:

Stephen J Thoma, Educational Psychology, The University of Alabama, 310 Carmichael Hall, Box 870238, Tuscaloosa, AL 35487, USA.

Email: [sthoma@ua.edu](mailto:sthoma@ua.edu)

## Introduction

The Defining Issues Test (DIT) was developed by James Rest in the 1970s to provide an alternative method and more user-friendly assessment of Kohlberg's model (Rest, 1979). Researchers who were more familiar with the Kohlberg interview approach were skeptical about the DIT's use of a more traditional paper-and-pencil format. To address these concerns, Rest began a careful process of developing a case for the measure that blended criteria from developmental theory, moral psychology, and traditional psychometrics. This research and development program has been ongoing for about 40 years and has generated sufficient empirical evidence to maintain the DIT as a central measure within moral psychology (Thoma, 2014). In our response to Curzer et al. (2014, hereafter Curzer), we draw off this work to highlight how DIT researchers have anticipated and addressed criticisms that have now resurfaced in the target article.

Paralleling the empirical work on the DIT, Rest and colleagues also periodically reevaluated the theoretical model that was used, in part, to underpin the measure. Although originally affiliated with Kohlberg's six-stage model, the Rest approach began to diverge and ultimately developed a theoretical position informed by, but distinct from, Kohlberg's theory. The resulting 'Neo-Kohlbergian' model embraces, modifies, and in some areas abandons portions of Kohlberg's developmental theory. These modifications have been well documented and presented in visible outlets. However, the fact that the model has evolved has led some who evaluate the DIT enterprise to make inferences based on either Kohlberg's views or earlier theoretical statements. As we will note later in this article, one central area in which the neo-Kohlbergian and orthodox Kohlbergian modes differ is in the proposed links between the psychological theory and philosophical positions. We will note that Curzer makes assumptions about the philosophical underpinnings of the DIT that are not part of DIT theory.

## The four-component model

Perhaps the best place to start our response is to highlight the working model of moral functioning that is known as Rest's four-component model (FCM; Rest, 1983, 1986). Developed in the 1980s as a response to Blasi's (1980) challenge to fill the judgment and action 'gap', the FCM was Rest's identification of psychological processes that ought to have some independent contribution to moral action. That is, the FCM suggests processes that partner with moral judgments in the formulation of moral action including moral sensitivity, moral judgment, moral motivation, and moral implementation. In the target paper, Curzer incorrectly describes the FCM as a model of moral judgments implying that it simply prised out the moral judgment process into subcomponents. On the contrary, the FCM was designed as a model of moral action and identifies the different components or semi-independent psychological processes that are marshaled to take action in a moral situation. Thus, moral sensitivity has its roots in social development and relies heavily on the individual's understanding of social phenomenon at the individual and group levels. Similarly, the moral motivation component subsumes processes that describe the individual's sense of self and the centrality of moral considerations within one's identity. Fundamentally, the FCM suggests that we must go beyond moral

judgment to explain moral action, and in so doing better represent how the individual brings multiple systems to bear on moral situations.

The eight criteria Curzer lists as problematic for the DIT can be clustered into two main concerns: (1) problems with the structure of the DIT and (2) that the DIT makes either inappropriate assumptions about the individual's moral theories or conflates these theories with moral judgments. We will address both clusters below.

## Questions about the structure of the DIT

It may be helpful to begin this section with a description of the DIT and how it is claimed to estimate an individual's moral judgment development. Within the FCM, the DIT is located in the moral judgment component (Component II). It is not the only measure of moral judgments identified by Neo-Kohlbergians nor is it viewed as representing the totality of Component II. Beginning in 1999 and the presentation of the Neo-Kohlbergian position, the DIT is identified as a measure of generalized moral schemas. These schemas were described as representing three developmentally ordered conceptions of moral understanding that capture how the individual views cooperation and associated conceptions of fairness. The model and empirical evidence suggests that the individual begins in early adolescence to view cooperation through a personal lens as evidenced by the attraction to items attending to how the individual or people known to the individual are affected by a particular moral issue. In middle adolescence, this conception gives way to a growing understanding of the moral underpinnings of society-wide norms, laws, and practices, which in turn is supplanted in late adolescence and emerging adulthood by a growing sense of the shared values that underlie social structures.

*How does the DIT work?* The DIT includes multiple stories (six in the original DIT-1 and five in DIT-2) in which the participant is presented with a moral dilemma and then asked to choose an action choice for the protagonist in the story. Following the choice, participants are asked to rate how important to their decision are 12 considerations, the majority of which provide only a hint of a rationale supporting the action choice. Mixed into these items are meaningless statements used to ensure that the participant is not attending to superficial cues and idiosyncratic response sets. After rating these statements, the measure requires participants to revisit all of the items associated with a story and rank the four most important considerations for their decision. The primary measures derived from the DIT focus on the ranking tasks. Note that it is not correct, as Curzer states, that the indices are derived simply from Likert scale summaries. Instead, the scores derived from DIT data come from the ranking data following the ratings and as such represent a second pass across the items and a consideration of the set of items as a whole. Note too that there are multiple scores derived from the DIT: an overall developmental index, scores on the three schemas, and scores representing features of the decision-making process. The characterization of the DIT presented by Curzer neglects to mention these features and outcome measures.

Since 1999 (Rest et al., 1999b), the DIT is said to be a schema activation measure. The claim that the DIT activates schemas follows from the use of sentence fragments. As typically described, schemas are knowledge structures that help the individual make inferences about incoming information in order to quickly understand the

situation and prepare to act in an organized way. DIT sentence fragments mimic this real-life process by providing just enough information to trigger the moral schema. What is theoretically assumed and empirically supported by significant evidence both within the DIT literature and more broadly in the developmental sciences is that a schema triggered by the story provides the scaffolding for interpreting the situation in more depth. By extension, the interpretation that is most important to the individual represents his or her preferred schema (Narvaez and Bock, 2002). To counter spurious responses, the measure repeatedly instructs the participant to give low ratings to ideas that are not readily understood and notes that these items should not be ranked. The sprinkling of meaningless items throughout the test encourages this test-taking set. It was established early in the DIT enterprise that participants do not randomly select items, that the rankings of these items reflect their current developmental moral understandings, and that with age and education these item preferences shift toward the more elaborate moral perspectives (Rest, 1979).

Curzer asserts without any empirical evidence that the DIT has four main structural problems. First, as was foreshadowed above, he claims that the sentence fragment approach is prone to elicit any number of interpretive systems rather than targeted moral understanding. Of course the concern Curzer raises is central to the development of any measurement instrument and is typically understood in terms of reliability and validity. If, as Curzer suggests, DIT items are as likely to evoke idiosyncratic responses as moral schema, then we would expect to find poor internal reliability estimates as there would be nothing coherent about individuals' responses. This is not the case. Similarly, if the DIT's ability to assess moral schemas were distorted by idiosyncratic responses, one would be hard-pressed to demonstrate links to validating criteria. Instead, the reliability and validity data supporting the DIT are substantial and support the measure as reliable, developmental, and within the moral domain (to explore the seven clusters of studies supporting the DIT as a valid measure of moral judgment development, see Rest, 1979, 1982, 1986; Rest et al., 1997, 1999a,c; Thoma, 2006, 2014; Thoma et al., 1999a,b).

Second and following from the criticism that the DIT elicits idiosyncratic responses, Curzer goes on to assert that the dilemma coverage is narrow and unrepresentative of the range of potential ethical dilemmas one might encounter resulting in unrepresentative estimates of one's reasoning. Certainly, no measure covers every possible scenario. But as mentioned previously, the DIT provides estimates of judgments that are broad-based and generalized, not ones concerned with attending to particular contexts. Instead, dilemmas are used to trigger the schema, and then responses indicate which defining issues are prioritized. At one point in the development of the DIT-2, we asked participants to consider up to 11 dilemmas in order to see how much estimates of moral judgment development changed (Rest et al., 1997). What we found was that soon after five dilemmas, the increased precision of the statistical estimates began to decline and the degree of improvement in the scores did not justify the added participant time and effort. In short, the DIT provides a snapshot measure of moral schemas and not a comprehensive measure of schema activation in every possible context.

Third, Curzer suggests, without empirical evidence, that development is too uneven across life's spheres to be measured by a single score on a few dilemmas. The criticism of limited coverage could be leveled at any measure and reflects the compromises

measurement developers must make to provide stable estimates while being mindful of the limits of participants' energy and attention. However, it is not accurate to suggest that the DIT boils down moral judgment into a single score. As DIT research developed, more and more scores were created from new analyses. Computer-generated scoring now provides a range of indicators of developmental level, a profile of schema usage, the utility of schema information in making moral decisions, and estimates of whether one is developmentally in transition or consolidated on a particular schema, among others (c.f. Thoma, 2006).

Although we have attempted to provide as much information as we can from DIT responses, we readily acknowledge that the DIT is not a comprehensive measure of moral development but only an aspect of it. It is a tool that provides a picture of developmental features of moral judgment. Similarly, we note that the FCM is an attempt to better understand the component processes that underlie moral development and we recognize that development may be uneven across these components. For instance, one may be able to recognize subtle features of moral situations without having similar strengths in moral judgments and so on. However, for what it measures, the DIT is stable over time, has reasonable internal consistency estimates, can predict action in different situations and across time, and can differentiate groups that ought to be different based on life experiences and training, among other validity criteria established by DIT researchers (Rest, 1979, 1986; Rest et al., 1999a; Thoma, 2014). To argue that the information the DIT produces fluctuates from situation to situation is hard to justify given the vast empirical evidence demonstrating quite the opposite.

Fourth, Curzer expresses concerns about the macro-morality focus of the DIT. That is, the measure is understood to assess a *broad* sense of how one views human cooperation rather than a more fine-grained estimate of the morality of everyday life. This distinction, although accurate, does not imply that the measure has nothing to say about the morality of human exchanges or micro-morality. Indeed, the DIT has been shown to be linked to micro-moral indicators such as behavior, moral interventions, and personal choices (Rest et al., 1999a). Additionally, and to supplement the information provided by the DIT, researchers in the FCM and Neo-Kohlbergian perspective have developed measurements that do provide the fine-grain assessment of moral judgments in specific settings (Bebeau and Thoma, 1999). Described as intermediate concept measures, these assessments are particularly well-suited to measure the moral judgments of individuals within particular contexts. Interestingly, researchers who have explored the relationship between the DIT and these more fine-grained, contextually relevant measures of moral thinking consistently find significant links between them (Thoma et al., 2008, 2013).

Curzer further zeroes in on the macro-moral level of assessment claiming that only a subset of individuals can coherently address macro-moral issues; the vast majority of participants using the DIT would be unaware of these issues. But this is a misunderstanding of how the term macro-morality is being used. In claiming that people are ignorant of macro-moral issues and thus fill out the DIT in an unreasoned way, Curzer mistakenly conflates macro-morality with knowledge about public policy that is known only to those who are familiar with a particular policy debate. Instead, as we view the term, macro-morality is concerned with an understanding of the basic structures of society including laws, social norms, institutions, and general practices common to the population and how

they help organize human cooperation (Rest et al., 1999b). The DIT asks participants to reflect on concrete situations as a window into how the individual prioritizes the dilemma's defining issues. For instance, in the now famous Heinz and the drug dilemma, the DIT-1 assesses the participant's focus on the protagonist, Heinz, whose dying wife needs a drug that Heinz cannot afford and, in a panic, contemplates stealing. From the pattern of participant responses, we can assess whether respondents focus on whether Heinz has the right motivation, the importance of laws and norms, or whether respondents prioritize the shared values evoked by the story. These are all questions that have been shown to be readily accessible to adolescents and adults as evidenced by the original design studies (Rest, 1979) and by subsequent tracking of the psychometric properties of the measure (Thoma, 2014). Note that the DIT does not ask the participant to reflect on poverty, health care, and crime as public policy concerns, all of which may be evoked by the Heinz dilemma but are not central to the DIT's mission or its summary scores.

## Questions about the underlying theory presupposed by the DIT

In addition to the concerns about how the DIT works and the information it provides, Curzer raises a number of questions concerning the link between moral theories and DIT scores. In this discussion, it is somewhat unclear to us what counts as a moral theory. In some instances, the term is used as a stand-in for moral schema and represents the individual's interpretation of moral phenomena from a psychological perspective. In other uses, the term moral theory is tied to formal philosophical perspectives. In general, Curzer's claim is that the DIT suffers from mixing an individual's moral theory with moral judgments, leaving the resulting estimates uninterpretable. To explore these criticisms, we acknowledge that the relationship between ethical theories and moral judgment development has a long history, particularly in the Kohlbergian model where associations between the psychological and philosophical models were encouraged. In particular, Kohlberg (1981) was clear in the assumption that developmentally advanced positions in the stage sequence were aligned with foundational principlism generally and John Rawls' views in particular. Although when conceived during the 1960s and 1970s, Kohlberg's adoption of Rawlsian notions may have been a prudent and successful strategy, we agree that changes both within psychology and philosophy make such claims increasingly untenable. As the neo-Kohlbergian approach makes clear, Kohlberg's claims about foundational principlism were counterproductive and unnecessary. Instead, a 'big-tent' view was adopted, suggesting that a post-conventional schema can be understood from a variety of philosophical traditions as long as they provide a mechanism for organizing cooperation in society that is sharable with all, can be debated, is logically consistent, and is coherent with acceptable practice (Rest et al., 1999b). Curzer echoed earlier criticisms of Kohlberg's theory by suggesting that the DIT's three developmentally ordered schemas represent a presumed hierarchy of philosophical moral theories, and that partisan claims were being made about which theories are superior to others. On the contrary, the aforementioned points about the link between moral theories and moral judgment development make clear that this claim is untrue both theoretically and empirically – we are unaware of any empirical evidence to suggest that the three developmentally ordered schemas align

with different philosophical moral theories. Note too that the developmental ordering of schemas is supported by psychological data and never attributed to philosophical moral theories (e.g. Rest et al., 1999b). That is, three developmentally ordered schemas are suggested because they are consistently observed in sample after sample of data, and are shown to develop in the proscribed order (e.g. Thoma, 2014).

Curzer extends the point about connections between philosophical moral theories and moral schemas by suggesting that the DIT fails because of the lack of consensus among philosophers about what one should do in the different DIT dilemmas. We would note that the choice of what the protagonist ought to do is not central to the workings of the DIT. That is, the DIT measures which of the three schemas are elicited by the dilemma – how the individual defines the dilemma – and is relatively uninterested in the actual choice about what the protagonist ought to do. For instance, the scoring of the DIT never attends to the action choice, except to note consistency between actions and reasons, but actions do not figure into the schema designations (Rest et al., 1997).

## Conclusion

The primary claims about the DIT advanced by Curzer cannot be sustained because of their failure to engage the literature supporting the measure. As we noted above, not once is this literature addressed or seriously considered resulting in statements and claims about the DIT that are easily dismissed because they are counter to the existing evidence. If this literature were consulted, it would be evident that researchers within the neo-Kohlbergian tradition have been careful to define the limits of measures designed to represent aspects of moral functioning. Regarding the DIT, we are well aware that it does not measure the totality of moral development, and a cursory review of the DIT literature will establish that we have never claimed it did. What the DIT does measure, it does with the precision expected for a psychological assessment within the moral domain. Because great care was taken to establish its theoretical, developmental, and psychometric properties the DIT has survived and continues to be the most often used measure of moral judgment development in the field.

## Acknowledgements

The second and third authors contributed equally to the manuscript and are listed in the given order by a coin toss.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## References

- Bebeau M and Thoma SJ (1999) 'Intermediate' concepts and the connection to moral education. *Educational Psychology Review* 11: 343–360.
- Blasi A (1980) Bridging moral cognition and moral action: A critical review of the literature. *Psychological Bulletin* 88: 593–637.

- Curzer HJ, Sattler S, DuPree DG, et al. (2014) Do ethics classes teach ethics? *Theory and Research in Education* 12(3): 366–382.
- Kohlberg L (1981) *Essays on Moral Development: The Philosophy of Moral Development: Moral Stages and the Idea of Justice*, vol. 1. San Francisco, CA: Harper & Row.
- Narvaez D and Bock T (2002) Moral schemas and tacit judgement or how the Defining Issues Test is supported by cognitive science. *Journal of Moral Education* 31(3): 297–314.
- Rest JR (1979) *Developing in Judging Moral Issues*. Minneapolis, MN: University of Minnesota Press.
- Rest JR (1982) A psychologist looks at the teaching of ethics: Moral development and moral education. *The Hastings Center Report* 12(1): 29–36.
- Rest JR (1983) Morality. In: P Mussen (Ed) *Carmichael's manual of child psychology*, fourth edition, volume on cognitive development, J Flavell and E Markman (Volume Eds). New York, NY: Wiley, pp. 556–629.
- Rest JR (1986) *Moral Development: Advances in Research and Theory*. New York: Praeger.
- Rest JR, Narvaez D, Bebeau M, et al. (1999a) A neo-Kohlbergian approach to moral judgment: An overview of Defining Issues Test research. *Educational Psychology Review* 11(4): 291–324.
- Rest JR, Narvaez D, Bebeau M, et al. (1999b) *Postconventional Moral Thinking: A Neo-Kohlbergian Approach*. Mahwah, NJ: Erlbaum.
- Rest JR, Narvaez D, Thoma SJ, et al. (1999c) DIT2: Devising and testing a new instrument of moral judgment. *Journal of Educational Psychology* 91(4): 644–659.
- Rest J, Thoma S, Narvaez D, et al. (1997) Alchemy and beyond: Indexing the Defining Issues Test. *Journal of Educational Psychology* 89(3): 498–507.
- Thoma SJ (2006) Research using the Defining Issues Test. In: M Killen and JS Smetana (eds) *Handbook of Moral Psychology*, pp. 67–92. Mahwah, NJ: Erlbaum.
- Thoma SJ (2014) Measuring moral thinking from a neo-Kohlbergian perspective. *Theory and Research in Education* 12(3): 347–365.
- Thoma SJ, Bebeau M and Bowland A (2008) The role of moral judgment in context-specific professional decision-making. In: F Oser and W Veugelers (eds) *Moral Development, Character and Action*. Rotterdam: Sense Publishers, pp. 147–160.
- Thoma SJ, Barnett R, Rest J, et al. (1999a) What does the DIT measure? *British Journal of Social Psychology* 38(1): 103–111.
- Thoma SJ, Derryberry P and Crowson HM (2013) Describing and testing an intermediate concept measure of adolescent moral thinking. *European Journal of Developmental Psychology* 10: 239–252.
- Thoma SJ, Narvaez D, Rest JR, et al. (1999b) Does moral judgment reduce to political attitude or verbal ability? *Educational Psychology Review* 11(4): 324–342.

## Author biographies

Stephen J Thoma is a University 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 is a Fellow of the American Educational Research Association and the American Psychological Association.

Muriel J Bebeau is Professor in the Department of Primary Dental Care, School of Dentistry; Faculty Affiliate in the Center for Bioethics; and Adjunct Professor in the Department of Educational Psychology at the University of Minnesota where she also, until 2008, directed the Center for the Study of Ethical Development. She pioneered an interdisciplinary, evidence-based



approach to teaching of ethics in dentistry, designing, and validating measures of functional processes that give rise to morality. She is an honorary fellow in the American College of Dentists and a Fellow in the American Educational Research Association.

Darcia Narvaez is Professor of Psychology at the University of Notre Dame. She publishes extensively on moral development and education. Her most recent book is *Neurobiology and the Development of Human Morality: Evolution, Culture and Wisdom* (2014, Norton). She is a fellow of the American Psychological Association. She is an Executive Editor of the *Journal of Moral Education*. She also writes a popular blog for *Psychology Today* ('Moral Landscapes').