

## Does the Defining Issues Test Measure Ethical Judgment Ability or Political Position?

CHARLES D. BAILEY  
*University of Memphis*

---

**ABSTRACT.** This article addresses the construct validity of the Defining Issues Test of ethical judgment (DIT/DIT-2). Alleging a political bias in the test, Emler and colleagues (1983, 1998, 1999, 2007), show that conservatives score higher when asked to fake as liberals, implying that they understand the reasoning associated with “higher” moral development but avoid items they see as liberally biased. DIT proponents challenge the internal validity of faking studies, advocating an explained-variance validation. This study takes a new approach: Adult participants complete the DIT-2, then evaluate the raw responses of others to discern political orientation and ethical development. Results show that individuals scoring higher on the DIT-2 rank others’ ethical judgment in a way consistent with DIT-2-based rankings. Accuracy at assessing political orientation, however, is low. Results support the DIT-2’s validity as a measure of ethical development, not an expression of political position.

**Keywords:** construct validity, defining issues test, moral judgment, political orientation

---

THE DEFINING ISSUES TEST of ethical judgment is a widely accepted instrument with a long track record; about 500 researchers use the DIT every year and have done so at a steady pace for the last 15 years.<sup>1</sup> Perhaps the greatest challenge to its construct validity is the claim that it is biased by political content—a claim traceable primarily to the study by Emler, Renwick and Malone (1983), which has stirred debate until the present (e.g., Crowson and DeBacker, 2008). Emler and his colleagues have based their challenge largely on test-takers’ ability to elevate their scores by “faking” as liberals; but proponents of the instrument have dismissed faking as a flawed research method and have relied on demonstrating that DIT scores explain variance over and above the test’s political content.

---

*Address correspondence to Charles D. Bailey, University of Memphis, Fogelman College of Business and Economics, School of Accountancy, 3665 Central Ave., Memphis, TN 38152, USA; cbailey2@memphis.edu (e-mail).*

This study takes a new approach: It empirically tests the “social communication” proposition offered by Emler and Stace (1999), which holds that political conservatives “do not obtain ‘lower’ scores on moral reasoning measures because they are incapable of obtaining ‘higher’ scores, but because the moral arguments they express convey the political identity they wish to communicate” (p. 458). In the current study, participants have personally taken the DIT attempt to evaluate the responses of other individuals, either as to their ethical judgment or to their political position.

### *Kohlberg’s Cognitive-Developmental Theory and the DIT*

Lawrence Kohlberg’s cognitive-developmental theory of ethical judgment has been the reigning paradigm in ethics-related research for the past half century. Throughout his extensive works, Kohlberg (e.g., 1969) argued that ethical judgment capacity in individuals progresses through six stages,<sup>2</sup> starting with rudimentary self-interest (promoting reward, avoiding punishment). Most adults reach the “Conventional” stage of reasoning (Rest, Narváez, Bebeau, & Thoma, 1999, p. 54), in which they centrally value the ideals of society, as expressed in laws, professional rules, and conventions. The highest Kohlbergian stage, however, is Postconventional, or Principled, reasoning, in which a person has internalized principles of justice or behavior that supersede laws or conventions when a conflict of values arises. To measure level of moral development, Kohlberg and his colleagues developed a labor-intensive Moral Judgment Interview approach (Colby & Kohlberg, 1987).

An influential breakthrough in the application of Kohlberg’s theory arose from the work of James Rest and colleagues at the Center for the Study of Ethical Development at the University of Minnesota, who developed paper-and-pencil instruments designed to measure ethical judgment ability along a developmental scale. The Defining Issues Test, or DIT (now in use for three decades), and DIT-2 (introduced around 1998, with contemporary ethical scenarios and slightly greater reliability, efficiency and validity) present ethical dilemmas and ask respondents to choose the ethical arguments that they consider most compelling.

An important assumption underlying the recognition-based responses used in the DIT—versus the self-produced ethical arguments used in Kohlberg’s Moral Judgment Interview—is that an individual will prefer arguments that represent the highest level of development they have attained but will not comprehend (and thus not find attractive) arguments based on higher levels of thinking than they personally have achieved. According to Rest and colleagues, a recognition test like the DIT can be as accurate in assessing judgmental capacity as would the self-produced arguments. The DIT instruments and their validation are extensively discussed in Rest et al. (1999).

The most-used DIT metric has been the *P* score, which is the percentage of the respondent’s chosen arguments that the developers contend represent the

Principled level of reasoning. Concurrent with introduction of the DIT-2, the Center for the Study of Ethical Development introduced an additional metric, *N2*, where the *N* denotes “new.” *N2* incorporates more information than *P*, but its interpretation is essentially the same. While *P* uses only subjects’ *rankings* of the importance of ethical decision factors, *N2* also considers the related *ratings* that respondents always have developed en route to stating their rankings. Thus *N2* scores can be computed for either the DIT-2 or the original DIT, which still is available.

It should be noted that the approach represented by the DIT does not adhere to a hard-stage theory, but recognizes that a person’s moral reasoning can reflect a range of “stages” at any point in their development. Rest and colleagues refer primarily to schemas rather than stages.

Following Kohlberg, DIT researchers accept the notion that an understanding of moral phenomena is developmental, and they agree with Kohlberg that moral judgment development begins with an early focus on the self and personal relationships, proceeds through an understanding of social conventions, and leads ultimately to a recognition of Postconventional (Principled) concepts. Accordingly, development is described in terms of three ordered schemas: Personal Interest, Conventional, and Postconventional. Extended discussions of the neo-Kohlbergian approach are found in Rest et al. (1999), the *Journal of Moral Education*’s special issue on the Minnesota approach to moral psychological research (Thoma, 2002), and Thoma (2006). Despite this divergence from Kohlberg, the distinction between Postconventional reasoning (favored more by liberals) and Conventional reasoning (favored more by conservatives) remains intact.

### *Challenges to the DIT Based on Political Content*

The popularity of the DIT has, appropriately, attracted challenges to its validity. Rest and colleagues (1999, Ch. 4) address six criteria applied to the validation of the DIT. Because of the rather high correlations in some research studies, they acknowledge that political content appears to represent the greatest threat to the construct validity of the DIT, and they devote nine pages (about 5% of the book) to discussing the correlation issue.

A well-documented correlation exists between conservative political ideology and Kohlbergian or DIT-based measures of moral reasoning (Alker & Poppen, 1973; Fishkin, Keniston, & MacKinnon, 1973; Candee, 1976; Johnson & Hogan, 1981; Lind, Sandberger, & Bargel, 1982; Nassi & Abramowitz, 1979; Rest, 1979). The DIT-proponents’ “orthodox” interpretation of the relationship is that philosophical or ideological *content* of reasoning is affected by moral stage or *structure* of reasoning. Rest and colleagues (1999) contend that political preferences naturally are highly correlated with moral judgment—implying, in view of the known correlations, that morally mature people tend to favor liberal

political ideology. Rest et al. (1999) “argue that DIT scores are especially illuminating of political decision making and behavior. How people respond to the old Kohlberg chestnuts (like ‘Heinz and the drug’) yields very useful information in understanding ‘red meat’ political controversies of the day (e.g., abortion, prayer in the schools, legalities in arresting and convicting the accused, rights of homosexuals, free speech, and women’s roles)” (p. 84).

Studies in the social psychology literature, however, notably Emler, Renwick, and Malone (1983) and Markoulis (1989), have argued that the orthodox interpretation is not correct. In these two studies, participants demonstrated the ability to alter their DIT scores by complying with instructions to respond to the DIT from a particular political perspective. The authors interpret their results as indicating that the DIT *P* score is more an artifact of political expression than a measure of ethical judgment capacity.

Emler and colleagues (1983) conducted an experiment in which participants were initially asked to respond to the DIT from their own perspective, and then were asked to generate “faked” responses, as a person of extreme conservative or extreme radical political persuasion. The participants’ own political persuasions were measured between the two administrations of the DIT. As the authors predicted, participants were able to fake their DIT scores successfully up as well as down. That is, conservative participants produced higher *P* scores when responding as a “liberal,” liberals produced lower scores when faking as “conservatives,” and moderate participants raised their scores if responding as a liberal and lowered their scores if responding as a conservative. While Kohlberg’s developmental theory would predict successful downward-faking, the authors argue it would predict that individuals could not fake upward, because they could not understand higher-level thought. If, however, the DIT items were reflecting only political orientation—which is not an established developmental characteristic—successfully faking upward would be a consistent result. The authors conclude that the study supports their claims that the so-called Conventional-Principled distinction is one of political content, not of differing structural complexity along a developmental hierarchy. Markoulis (1989) produced similar results using university students and graduates in a Greek (versus Scottish) environment.

To refute Emler et al. (1983), Barnett, Evans, and Rest (1995) conducted a similar faking experiment in which they modified the DIT to include 16 additional anti-establishment (A) items,<sup>3</sup> testing the alternative explanation that experimental demands, rather than true insights into liberal thought, drove participants to choose “Principled” items. They reasoned that participants in the original study, when attempting to respond as radicals, had endorsed “Principled” items simply because they had exhausted the available “A” items. Their results supported this hypothesis, but ironically their own study employed a modified test as well as modified instructions, contrary to the DIT proponents’ own arguments against

test-manipulation methods. For example, Thoma, Barnett, Rest, and Narváez (1999a) note that,

Classical measurement theory holds that the test items *and instructions* [emphasis in the original] are an integrated whole; changing the instructions is presumed to alter the test and what it measures. . . . Thus, it is no surprise that the manipulation of test instructions can produce variance in DIT scores, but this kind of study is insufficient to say how participants normally produce the variance of DIT scores. The manipulation invalidates the very process we want to study. (p. 108)

In a response to the rebuttal by Barnett and colleagues (1995), Emler, Palmer-Canton, and St. James (1998) report another series of four studies showing that arguments appearing in the DIT are interpreted as political statements. Further, they find that Conventional moral arguments carry a clearer political meaning than Principled moral arguments; and, since the *P* score is expressed as a *percentage* of most-favored arguments that are Principled, the ipsative scoring system may artificially reduce the *P* scores of conservative individuals. Specifically, since they view Conventional items as important to their political identity, they will embrace them; and the ipsative nature of scoring prevents scoring high on *both* Conventional and Principled reasoning (endorsing both strongly).

With particular relevance to the current study, Emler and Stace (1999) argue further for an alternative, “social communication” view . . . that these different forms of moral reasoning are expressions of contrasting political identities, and differ in ideological content rather than developmental level (p. 455). . . . [Conservatives] do not obtain “lower” scores on moral reasoning measures because they are incapable of obtaining “higher” scores, but because the moral arguments they express convey the political identity they wish to communicate. . . . The social communication view . . . *leads to the prediction that moral arguments used as self-presentations to communicate political identities will be appropriately interpreted in this way by their audience.* (pp. 458–459, emphasis added)

The current study directly tests this prediction based on the social communication view.

Emler, Tarry and St. James (2007) continue to explore the relationship between political orientation and this social-communication view by directly studying perceptions of persons by their peers within established social groups. They examine correlations between DIT responses (including *P* scores and ratings of item importance) and the individuals’ reputations, assuming that moral reasoning may have affected reputations in predictable ways. They find that, when Principled reasoning is measured by the *ratings* given to Principled items, the relationship to politics is not found, while based on the ipsative *P* score it is. With Conventional items, the relationship to politics is found with either rating- or ranking-based metrics. This supports the proposition that Conventional reasoning conveys a political identity, while Principled reasoning does not and is

equally understood and appreciated by persons at the so-called Conventional *and* Postconventional/Principled levels.

The “faking” studies raised challenging questions about the political content of the DIT. However, the internal-validity criticisms have led to an intellectual standoff concerning their interpretation. As noted above, Rest et al. (1999) acknowledge that political content appears to represent the greatest threat to construct validity, although they argue that a moderate correlation is normative. In defense of their instruments, they report extensive validation studies to show that the DIT explains significant additional variance over the other variables with which its scores are correlated. Thoma, Narváez, Rest, and Derryberry (1999b) argue for such a variance-based definition of validity. They review two dozen studies to address the discriminant validity of the DIT, as distinct from both political attitudes and verbal ability, and conclude, “Our findings are unambiguous: consistently, the DIT provides unique information above and beyond that accounted for by verbal ability, general ability, political attitudes, or political identity” (p. 338). Similarly, Crowson and DeBacker (2008) showed that DIT-2 scores contributed above and beyond (and quite independently from) political measures, in explaining students’ attitudes towards civil liberties and human rights.

### The Current Study

The experiment reported here avoids the faking device and directly tests the information content of the DIT-2 instrument. If the DIT-2 is a test of moral judgment, as intended, then persons scoring higher should have access to a broader spectrum of thought processes—i.e., they will comprehend moral arguments up to and including the developmental level they have achieved, while encountering fewer arguments that are incomprehensible to them. Accordingly, they should be able to rank the judgments of others more accurately as to their place on the developmental spectrum. Further, if the responses to the DIT-2 communicate ethical content more than political content, then the ability to classify persons politically should be less than the ability to rank them as to ethical judgment.

In this experiment, I examine the ability of a sample of adults to evaluate other target individuals as to (a) their ethical judgment level and (b) their political orientation, based on the targets’ responses to the DIT-2 instrument. Support for the validity of the DIT-2 (and presumptively the DIT) as a developmental measure of ethical judgment would emerge from results that support the following substantive hypotheses:

*H1:* Accuracy in judging ethical development from others’ responses to the DIT-2 instrument is positively related to the instrument’s rating of one’s own ethical development.<sup>4</sup>

*H2:* Accuracy in evaluating others’ political orientation from the DIT-2 is less than accuracy in evaluating others’ ethical development.

Conversely, failure of judges with higher DIT-2 scores to excel at evaluating the ethical development of others, combined with greater ability to discern political orientation, would be consistent with the arguments of Emler and colleagues (1983, 2007) and Emler and Stace (1999).

## Method

### *Participants*

Forty-four paid participants were recruited through a variety of campus announcements, advertisements in a campus newspaper, and networking. The objective was to recruit a diverse group of participants. All but two were U.S. citizens. Their ethical-judgment ( $N2$ ) scores ranged from very low ( $-1.4$ ) to high ( $65.4$ ), with a mean of  $39.2$ . This was similar to the target group that they would be evaluating ( $0.09$ ,  $63.2$ , and  $31.2$  respectively; see Table 1). Twenty-two were female and 22 male. Educational levels, based on the DIT-2 demographic scale, were college junior, 13; senior or undergraduate degree, 13; and professional or master's degree, 18. Ages ranged from 19 to 62, with a mean of  $29.4$ . Politically, again according to their responses on the DIT-2 demographic scale, 7 were very liberal, 10 somewhat liberal, 10 neither liberal nor conservative, 16 somewhat conservative, and 1 very conservative.

They received an average of \$62 for completing the experiment (which required about three hours), based on their judgmental accuracy in ranking and classifying target individuals, subject to a minimum of \$30.00 and a maximum of \$100.00.

### *Materials*

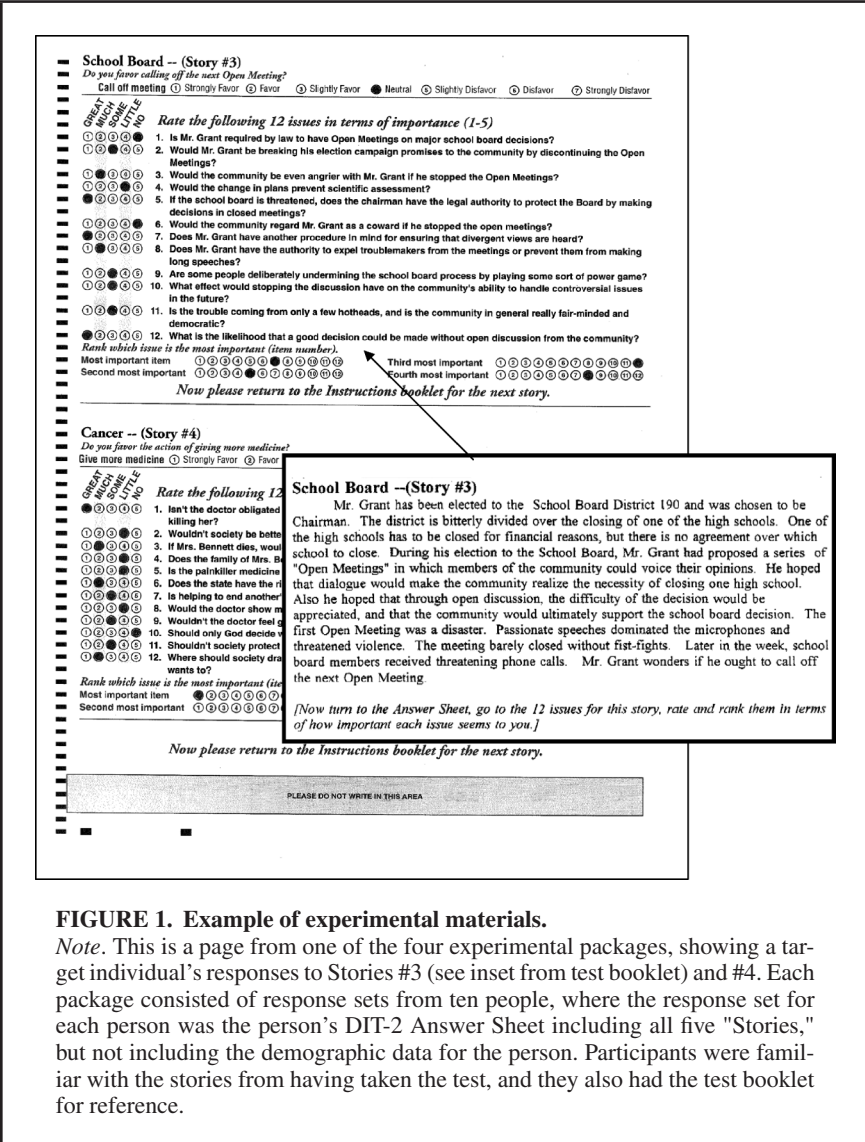
Participants completed the DIT-2 in the normal mode and subsequently examined photocopies of 20 target individuals' responses to assess the ethical judgments of 10 and the political orientations of another 10. Figure 1 shows one of the five scenarios with a target individual's responses marked. The sets of target individuals for evaluation were drawn from 82 DIT-2 instruments collected from previous participants in graduate business classes. Four sets of 10 instruments each were chosen such that, within each of the four sets, the key measure of Principled judgment,  $N2$ , varied in increments from a very low level to fairly high. The increments were kept at approximately  $1/3 SD$ , a level that represents a difference likely to be noticeable (Murphy & Myors, 2004, p. 13). The sequence of presentation within sets was randomized. Further, sets A, B, C, and D were staggered such that participants who received A for ethical evaluation received B for political evaluation, etc.<sup>5</sup> The combinations were randomly assigned to participants. The purpose of using four sets, rather than the minimal two sets required by the experimental design, was to provide some assurance against an internal

**TABLE 1. Descriptive Statistics**

Panel A: Ethical judgment (N2) scores and demographic data from the DIT-2						
	Mean	SD	Min.	Max		
Ethical judgment score						
Participant-judges ( <i>n</i> = 44)	39.2	14.9	-1.41	65.4		
Targets ( <i>n</i> = 40)	31.2	17.7	.09	63.2		
Age						
Participant-judges	29.4	10.9	19	62		
Targets	28.1	8.6	22	54		
Gender (% male/ % female)						
Participant-judges	50/50					
Targets	39/61					
Political orientation						
	Very liberal	Some liberal	Neither L/C	Some conservative	Very conservative	
Participant-judges	16%	23%	23%	36%	2%	
Targets	3%	38%	28%	20%	13%	
Educational level						
	College Junior	Senior or UG Degree	Professional or Master's degree			
Participant-Judges	29.5%	29.5%	41.0%			
Targets	0.0%	18.0%	82.0%			
Panel B: Dependent variables from the experiment (Kendall's Tau Statistic)						
	Mean	SD	Min.	Max	Lower 95% CI	Upper 95% CI
Accuracy in ranking as to ethical development ( $\tau_{eth}$ )						
All judges	.238**	.311	-.778	.733	.144	.333
Higher-N2 judges	.317**	.226	-.111	.733	.217	.417
Lower-N2 judges	.159*	.366	-.778	.733	-.003	.322
Accuracy in ranking as to professed political liberal/conservatism ( $\tau_{pol}$ )	.016	.246	-.511	.511	-.067	.099

Notes. \*\* $p < .001$ ; \* $p < .05$ , one-tailed;  $n = 22$  for higher- and lower-N2 groups, 44 total. N2 scores represent degree of principled (Postconventional) reasoning based on the DIT-2 instrument.





**FIGURE 1.** Example of experimental materials.

*Note.* This is a page from one of the four experimental packages, showing a target individual's responses to Stories #3 (see inset from test booklet) and #4. Each package consisted of response sets from ten people, where the response set for each person was the person's DIT-2 Answer Sheet including all five "Stories," but not including the demographic data for the person. Participants were familiar with the stories from having taken the test, and they also had the test booklet for reference.

validity threat (that the results might be driven by idiosyncratic responses in test materials) and enhance generalizability of the results.

Table 1, Panel A, shows demographic statistics and ethical judgment (*N2*) scores of the target individuals for comparison to the demographics of the judging participants. They are particularly similar in their ethical judgment scores and ages.

### *Design and Procedure*

The experiment consisted of three main phases, requiring about three hours in total: (1) personal completion of the DIT-2, (2) review of a package containing 10 persons' responses to this same DIT-2 test instrument (stripped of demographic data), to be ranked as to their levels of ethical judgment, and (3) review of a similar set from another 10 persons, to be classified according to the probable nature of the target individual's self-professed political orientation on the five-point scale incorporated in the DIT-2 instrument. Not every participant attended a single session combining all three phases of the experiment. Two participants, former students from whom I already had DIT-2 scores, received their materials by mail. Several attended two shorter sessions to suit their schedule, or completed the last phase at home. Consideration was given to avoiding fatigue and allowing participants ample opportunity to maximize their performance, and hence, their pay.

### *Data*

Valid DIT-2 scores were obtained from the 44 participants. As discussed above,  $N2$  is a new measure introduced with the DIT-2, which uses more data than the well-known  $P$  score but still reflects the proportion of Principled reasoning preferred by the respondent. Other measures of ethical reasoning are also provided from the DIT-2 and are relevant to this study; they are the  $P$  score and the average *importance ratings* assigned to items at Stages 2, 3, 4, 5, and 6.

The key experimental response variables were (a) the rankings, from highest to lowest, of the 10 target individuals' levels of ethical judgment and (b) the categorizations of the other 10 target individuals according to how they predict those individuals responded on the five-point DIT-2 scale: Very Liberal, Somewhat Liberal, Neither Liberal nor Conservative, Somewhat Conservative, and Very Conservative.

### *Measures of Accuracy in Ranking and Classifying Target Individuals*

Hypothesis 1 addresses the relationship between participants' DIT-2 ethical-judgment scores ( $N2$  or alternate measures of ethical development) and participants' accuracy in evaluating others on such scores. The experiment operationalizes accuracy as a correlation between the judge's subjective ranking of 10 target individuals and the ranking by actual DIT scores of the targets. Hypothesis 2 concerns the accuracy of evaluating political orientation. That relationship is operationalized as the correlation between judges' subjective evaluations of 10 targets along the five-point DIT-2 political scale and the target person's actual response on that scale. Since both "accuracy" metrics are rank-level measures, the nonparametric statistic Kendall's tau ( $\tau$ ) is appropriate. Given samples of intermediate size,  $\tau$  approaches normality more rapidly than does Spearman's  $r_s$  and

can be interpreted as an estimation of the population parameter, while  $r_s$  has no such corresponding population parameter (Daniel, 1978, p. 314). Tau also deals well with large numbers of tied rankings, which exist when assigning 10 persons to five political categories (Field, 2005, p. 131). Thus,

$\tau_{\text{eth}}$  = judge's accuracy in ranking ten targets as to ethical judgment ability, and  
 $\tau_{\text{pol}}$  = judge's accuracy in ranking ten targets as to political position.

## Results

Table 1, Panel B, shows the descriptive statistics for accuracy in ranking target individuals according to ethical development and classifying them according to political liberalism/ conservatism. The pattern is as hypothesized, with accuracy in ranking by ethics being positive and significant, while accuracy in political classification is essentially nil. Further, those scoring higher themselves on ethical judgment are able to rank others more accurately. Because the hypotheses are directional,  $p$  values stated below are one-tailed unless otherwise specified.

The accuracy in evaluating ethical development ( $\tau_{\text{eth}}$ ) is moderate in effect size and significantly positive. For the more "ethically mature" judges (dichotomized at the median based on  $N2$  scores), mean  $\tau_{\text{eth}} = .317$ ,  $t(21) = 6.59$ ,  $p < .001$ ,  $d = 1.41$ . For the less ethically mature judges, mean  $\tau_{\text{eth}} = .159$ ,  $t(21) = 2.05$ ,  $p < .05$ ,  $d = .43$ . Overall, mean  $\tau_{\text{eth}} = .238$ ,  $t(43) = 5.09$ ,  $p < .001$ ,  $d = .77$ . Mean accuracy in discerning political position, however, is only  $.016$ ,  $SD = .246$ ; 95% confidence interval between  $-.067$  and  $.099$ . Thus, the upper confidence interval indicates 97.5% confidence that the correlation does not exceed  $.099$ , a small effect size at most.

Moreover, the correlation between accuracy of ethical ratings ( $\tau_{\text{eth}}$ ) and the raters' own  $N2$  scores reveals a significant positive relationship ( $r[44] = .416$ ,  $p = .002$ ), supporting Hypothesis 1. This moderate-to-large effect size indicates that the participants have some insight into (and agreement with) the relative ethical content of the DIT-2 responses.<sup>6</sup>

Correlations between classification accuracy and other DIT-2-based measures of ethical development are even more illuminating. Table 2 shows these correlation coefficients and related significance using, as alternatives to  $N2$ , the average importance the judge had assigned to DIT-2 items related to Stage 2, Stage 3, etc. The correlation based on Stage 5 (one category of Principled reasoning) is highest ( $r[44] = .433$ ,  $p = .002$ ), but the correlation using Stage 2 also is significant and negative ( $-.294$ ,  $p = .027$ ), implying that individuals embracing this low level of judgment are below average at ranking others.<sup>7</sup> Endorsements of items from Stages 3 and 4 show no significant correlation with ranking accuracy, apparently because of their midrange status.

Accuracy of political categorization is, as noted above, quite low. Thirty-six participants completed the final phase of the experiment, judging ten other

**TABLE 2. Correlations Between Accuracy in Classifying Others' Ethical Judgment and Measures of the Judge's Ethical Development**

<i>Metric from the DIT-2</i>	<i>Correlation (r)</i>
Rating of Stage-2 items	-.294*
Rating of Stage-3 items	-.094 n.s.
Rating of Stage-4 items	-.059 n.s.
Rating of Stage-5 items	.433**
Rating Stages 5 & 6 items	.377**
N2 score	.416**
P score (percentage) <sup>a</sup>	.281*

\*\* $p < .01$ , \* $p < .05$ , one-tailed;  $n = 44$ .

Kohlbergian stages 2 and 3 items reflect the lowest, personal-interest schema of ethical reasoning; stage 4 reflects the "maintaining norms," or conventional schema; and stages 5 and 6 reflect a Postconventional, or principled, schema. (Rest et al., 1999, p. 54).

<sup>a</sup>For consistency, this correlation employs the judge's *P* score and the targets' *P* scores, not *N2*.

target individuals as to their expected response on the DIT-2's five-point political-orientation scale.<sup>8</sup> This is significantly less than the mean correlation ( $\tau_{\text{eth}}$ ) of .269 ( $SD = .258$ ) achieved by these 36 participants in evaluating ethical judgment (paired  $t [35] = 4.22$ ,  $p < .001$ ,  $d = 1.00$ ). Table 3 summarizes the judgments of political orientations versus the targets' self-professed orientations and illustrates the low level of agreement. The correct response categories are underlined, and the modal responses of the judges are bolded. In no instance is the modal choice correct, and the lack of differentiation is evident, with estimates ranging widely. Thus, substantive Hypothesis 2 is supported—accuracy in evaluating others' political orientation is less than accuracy in evaluating others' ethical development—and, somewhat surprisingly, little success at discerning political position is detectable.

## Discussion

This study contributes to the debate over the political content of the DIT/DIT-2. The results concerning the ability of lay persons to rank their peers' ethical judgment ability based on raw DIT-2 responses—and to do so more accurately if they themselves are at a higher "developmental" level according to the same instrument—may be more important than demonstrating the low ability to discern political position from DIT-2 responses. The overall findings lend support to the validity of the DIT-2 as a measure of ethical judgment ability.

The participants' task of assessing a target's ethical development by reading the target's responses, with no computational aid, seems clearly a difficult one.

**TABLE 3. Judged Political Orientation vs. Professed Orientation**

Target individual's professed political orientation	<i>Judge's prediction of political orientation</i>					Total
	Very liberal	Somewhat liberal	Neither liberal nor conservative	Somewhat conservative	Very conservative	
Panel A:			Cell Counts			
Very liberal (1)	0	0	1	7	0	8
Somewhat liberal (15)	12	35	32	47	8	134
Neither liberal nor conservative (12)	10	29	23	35	12	109
Somewhat conservative (7)	4	15	17	13	14	63
Very conservative (5)	5	12	10	14	5	46
Total evaluations	31	91	83	116	39	360
Panel B:			Percentages			
Very liberal	0%	0%	13%	88%	0%	100%
Somewhat liberal	9%	26%	24%	35%	6%	100%
Neither liberal nor conservative	9%	27%	21%	32%	11%	100%
Somewhat conservative	6%	24%	27%	21%	22%	100%
Very conservative	11%	26%	22%	30%	11%	100%
Total evaluations	9%	25%	23%	32%	11%	100%

*Note.* The five political categories are from the DIT-2 demographic data. Correct responses are underlined and modal responses are **bolded**. Numbers in parentheses are the target individuals who professed each orientation; with the four packages of ten, the total targets are 40. The total number evaluated in each category is a function of the number of such persons distributed throughout the four experimental packages and the numbers of packages returned.

Thus, one would not expect high accuracy from an untrained participant's impressionistic reading of the responses. In this light, the correlations between judged rankings and rankings based on the DIT-2 scores are surprisingly high. Moreover, the ability of judges' *importance ratings* of Principled/ Postconventional items (Table 2)—not just their ipsative *P* or *N2* scores—to predict their ability to evaluate their peers adds to the credibility of these reasoning stages as developmental levels.

It had seemed plausible that, to paraphrase Emler and Stace (1999, p. 459), *political identities of respondents would be appropriately interpreted by the audience*—even if one questions whether the communication is motivated. Thus, the finding that participants could gain so little insight into the target individuals' political orientation was unexpected, given the substantial monetary incentive and the established link between DIT scores and political conservatism.<sup>9</sup> Emler and Stace (1999) suggest that persons with more developed political positions may be more skillful at interpreting political signals or evaluating others' political orientation. The participants in the current study were adults, and 10 of the 36 who made the political evaluations held a master's or professional degree. Only nine (25%) indicated low political commitment by declaring themselves "neither liberal nor conservative."

It is not appropriate, nor is it my intention, to perpetuate an absolutist position about moral development and political attitudes. While Emler and his colleagues appear to claim that DIT scores are wholly a function of political orientation, the current study might be construed as evidence that DIT scores have nothing to do with political orientation. But these either/or positions are inconsistent with past research. Johnson and Hogan (1981) argue explicitly that political and ethical attitudes are inextricably intertwined, and that malleability of ethical-judgment scores when presenting one's self to a liberal or conservative audience is exactly what we would expect, not evidence against the idea that the instrument in question measures moral orientation. Similarly, Rest and his colleagues have consistently said that a correlation between ethical development and political orientation is expected (e.g., Rest et al., 1999). Thus, the results of this study serve only to counterbalance arguments at the other extreme.

### *Limitations*

The study cannot, of course, prove the absence of social-communication *motives* in the responses. Another fruitful approach to the question might be to examine the impact of an individuals' disposition towards "socially desirable responding" (e.g., Paulhus, 2002). If the "social communication" proposition of Emler and colleagues is correct, then persons high on such a tendency might respond to the DIT/DIT-2 in ways that reveal more about their political orientation than about their ethical judgment.

The measure of political orientation used here (from the DIT-2) does not discriminate between the two-factor representation cited by Emler et al. (2007, p. 87). Nevertheless, the DIT-2 scale concerning political views appears (appropriately for this discussion) to represent “the left-right ideological dimension, related to political party support and voting preferences but not to education” rather than the second factor, expressed as either “libertarian/authoritarian” or “democratic enlightenment.”

### *Conclusions*

The results of this study complement the evidence from studies by Thoma and colleagues (1999b) and Crowson and DeBacker (2008), which adopt an explained-variance approach. Instead, it directly tests the “social communication” view expressed by Emler and Stace (1999) and finds that the DIT-2 communicates evidence of ethical judgment ability that is interpreted more accurately by persons who themselves score higher on the test (and not as an artifact of rating by similarity to oneself). Ability to discern political orientation from the responses, however, is nil.

### NOTES

1. S. Thoma, Center for the Study of Ethical Development, in recent personal correspondence.
2. Briefly, the six Kohlbergian stages are as follows (Rest, 1979, p. 9):

Stage 1—Punishment and obedience orientation

Stage 2—Naïve instrumental hedonism.

Stage 3—Good-boy/girl morality of maintaining good relations, approval of others

Stage 4—Authority-maintaining morality

Stage 5—Morality of contract, of individual rights, and of democratically accepted law.

Stage 6—Morality of individual principles of conscience.

Stages 1–3 are characterized as Self-Interest, Stage 4 as Conventional, and Stages 5–6 as Principled.

3. The standard DIT contains five such A items, making available an additional Antisocial score that is not used in computing the ethical-judgment metrics. The Antisocial score represents a point of view critical of the Conventional order, but offering nothing constructive in its place (Bebeau & Thoma, 2003).

4. The reader may observe that an alternative phenomenon could account for such a correlation and must be ruled out: if judges evaluated targets whose responses agreed with theirs as having superior judgment, then the logic would be circular.

5. The option of evaluating the same target set on both criteria, while requiring less labor from participants, entails a threat of confounding. For example, after categorizing a target as having low ethical judgment ability, one would likely resist attributing their own political beliefs to that person.

6. An alternative explanation for this positive relationship between the rater’s own putative “ethical development” score and greater accuracy in predicting others’ level of “ethical development” is that raters may use *similarity to their own responses* as the benchmark for high ethical reasoning. Thus, high-scoring judges would tend to “accurately” assign high scores to similar target individuals, but low-scoring judges would “inaccurately” assign high scores to the targets similar to themselves. If so, then the logic of this study would be circular. To test that possibility, I generated, for each judge, a “similarity” measure between the judge and each of the ten targets he or she evaluated as to ethical

judgment. The DIT-2 instrument includes five ethical scenarios, each with 12 factors to be rated as to their importance in the decision, a total of 60 items. The similarity metric is a Pearson correlation between the ratings by the judge and by the target individual. Each judge evaluated ten target individuals; across all 44 judges, the average correlation between a judge's agreement with a target and his or her ranking of the target was  $r(10) = .18$  (N.S.). Moreover, a strategy of assigning higher ranks to targets who gave responses similar to the judge does not appear to have been effective; a correlation between the degree to which a judge's rankings were correlated with similarity-to-target, and the accuracy achieved, is *negative* and not significant ( $r[44] = -.18, p = .24$ ).

7. The description of Stage-2 thinking in Rest (1979, p. 22) illustrates why one operating at this low level would have difficulty understanding *either* conventional *or* principled reasoning: "although each person is understood to have his own interests, an exchange of favors might be mutually decided. . . . The [central concept determining rights and responsibilities is the] morality of instrumental egoism and simple exchange: 'Let's make a deal.'"

8. Eight of the 44 participants who completed the evaluations of ethical judgment did not complete the political evaluations. Deleting the eight participants from the analysis of ethical ratings does not change the conclusions.

9. Another indication why the judges might be expected to garner insights into the targets' political positions is found in a study quite outside the Kohlbergian paradigm, because it is purely empirical and does not rank moral values according to any system such as Kohlberg's ethic of justice. Graham, Haidt, and Nosek (2009) identify five sets of widely-shared "moral intuitions": Harm/care, Fairness/reciprocity, Ingroup/loyalty, Authority/respect, and Purity/sanctity. The first two closely resemble Kohlberg's Principled thinking, while the others represent lower levels within the DIT. They find that liberals place more importance on the first two, while conservatives place approximately equal emphasis on all five. Information about such relative preferences is available from the targets' raw DIT responses.

## AUTHOR NOTE

**Charles D. Bailey** is a Professor of accountancy and Arthur Andersen Chair of Excellence, University of Memphis.

## REFERENCES

- Alker, H. A., & P. J. Poppen. (1973). Personality and ideology in university students. *Journal of Personality, 41*, 652-671.
- Barnett, R. C., J. Evans, & J. Rest. (1995). Faking moral judgment on the Defining Issues Test. *British Journal of Social Psychology, 34*, 267-278.
- Bebeau, M. J., & Thoma, S. J. (2003). *Guide for the DIT-2*. Minneapolis, MN: Center for the Study of Ethical Development, University of Minnesota.
- Candee, D. (1976). Structure and choice in moral reasoning. *Journal of Personality and Social Psychology, 34*, 1293-1301.
- Colby, A., & Kohlberg, L. (1987). *The measurement of moral judgment*. New York, NY: Cambridge University Press.
- Crowson, H. M., & DeBacker, T. K. (2008). Political identification and the Defining Issues Test: Reevaluating an old hypothesis. *The Journal of Social Psychology, 148*, 43-60.
- Daniel, W. W. (1978). *Applied nonparametric statistics*. Boston, MA: Houghton Mifflin.
- Emler, N., Renwick, S., & Malone, B. (1983). The relationship between moral reasoning and political orientation. *Journal of Personality and Social Psychology, 45*, 1073-1080.
- Emler, N., Palmer-Canton, E. & St. James, A. (1998). Politics, moral reasoning and the Defining Issues Test: A reply to Barnett et al. (1995). *British Journal of Social Psychology, 37*, 457-476.



- Emler, N., & Stace, K. (1999). What does principled versus conventional moral reasoning convey to others about the politics and psychology of the reasoner? *European Journal of Social Psychology*, 29, 455–468.
- Emler, N., Tarry, H., & St. James, A. (2007). Principled moral reasoning and reputation. *Journal of Research in Personality*, 41, 76–89.
- Field, A. (2005) *Discovering statistics using SPSS*. (2nd ed.) London, UK: Sage.
- Fishkin, J., Keniston, K., & MacKinnon, C. (1973). Moral reasoning and political ideology. *Journal of Personality and Social Psychology*, 27, 109–119.
- Graham, J. D., Haidt, J. J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96, 1029–1046.
- Johnson, J. A., & Hogan, R. (1981). Moral judgments and self-presentations. *Journal of Research in Personality*, 15, 57–63.
- Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347–480). Chicago, IL: Rand McNally.
- Lind, G., Sandberger, J. & Bargel, T. (1982). Moral judgment, ego strength and democratic orientation. *Political Psychology*, 2, 70–110.
- Markoulis, D. (1989). Political involvement and socio-moral reasoning: Testing Emler's interpretation. *British Journal of Social Psychology*, 28, 203–212.
- Murphy, K. & Myers, B. (2004). *Statistical power analysis*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Nassi, A., & Abramowitz, S. (1979). Transition or transformation? Personal and political development of former Berkeley Free Speech Movement activists. *Journal of Youth and Adolescence*, 8, 21–35.
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 67–88). Hillsdale, NJ: Lawrence Erlbaum.
- Rest, J. (1979). *Development in judging moral issues*. Minneapolis, MN: University of Minnesota Press.
- Rest, J., Narváez, D., Bebeau, M., & Thoma, S. (1999). *Postconventional moral thinking: A neo-Kohlbergian approach*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Thoma, S. J. (Ed.). (2002). The Minnesota approach to moral psychological research [Special issue]. *Journal of Moral Education*, 31(3).
- Thoma, S. J. (2006). Research using the Defining Issues Test. In M. Killen and J. Smetana (Eds.), *Handbook of Moral Development* (pp. 67–91). Mahwah, NJ: Lawrence Erlbaum.
- Thoma, S., Barnett, R., Rest, J., & Narváez, D. (1999a). What does the DIT measure? *British Journal of Social Psychology*, 38, 103–111.
- Thoma, S. J., Narváez, D., Rest J., & Derryberry, P. (1999b). Does moral judgment development reduce to political attitudes or verbal ability? *Educational Psychology Review*, 11, 325–342.

Received August 14, 2009

Accepted March 12, 2010

Copyright of Journal of Social Psychology is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.