

This article was downloaded by: [University of Alabama at Tuscaloosa],
[Stephen Thoma]
On: 19 March 2013, At: 11:34
Publisher: Routledge
Informa Ltd Registered in England and Wales Registered Number: 1072954
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH,
UK



European Journal of Developmental Psychology

Publication details, including instructions for authors
and subscription information:

<http://www.tandfonline.com/loi/pedp20>

Describing and testing an intermediate concept measure of adolescent moral thinking

Stephen Thoma^a, W. Pitt Derryberry^b & H. Michael
Crowson^c

^a Educational Psychology Department, University of
Alabama, Tuscaloosa, AL, USA

^b Psychology Department, Western Kentucky
University, Bowling Green, KY, USA

^c Educational Psychology Department, University of
Oklahoma, Norman, OK, USA

Version of record first published: 07 Mar 2013.

To cite this article: Stephen Thoma, W. Pitt Derryberry & H. Michael Crowson (2013):
Describing and testing an intermediate concept measure of adolescent moral thinking,
European Journal of Developmental Psychology, 10:2, 239-252

To link to this article: <http://dx.doi.org/10.1080/17405629.2013.766129>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes.
Any substantial or systematic reproduction, redistribution, reselling, loan, sub-
licensing, systematic supply, or distribution in any form to anyone is expressly
forbidden.

The publisher does not give any warranty express or implied or make any
representation that the contents will be complete or accurate or up to

date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Describing and testing an intermediate concept measure of adolescent moral thinking

Stephen Thoma¹, W. Pitt Derryberry², and
H. Michael Crowson³

¹Educational Psychology Department, University of Alabama, Tuscaloosa,
AL, USA

²Psychology Department, Western Kentucky University, Bowling Green, KY, USA

³Educational Psychology Department, University of Oklahoma, Norman, OK, USA

This paper describes the development and preliminary testing of an intermediate concept measure (ICM) of moral thinking for adolescent populations. First proposed by Rest and Narvaez (1994), intermediate concepts are described as more context specific than moral stages defined within the Kohlberg tradition, but are more abstract than assessments of codes of conduct. The process of developing the adolescent ICM is described and data are presented to provide initial support for the measure. Results indicate that the adolescent ICM can distinguish age educational groups across high school and individuals who are acting out in school achieved significantly lower scores than all other students. In addition, ICM scores are related to Defining Issues Test scores providing preliminary support for the claim that both measures assess the moral domain. Coupled with the respectable psychometric properties of the measure, these findings support the adolescent ICM as a measure of moral thinking in adolescent populations and as a potential outcome measure for character education programmes.

Keywords: Moral judgement development; Measurement; Adolescence.

Researchers in the neo-Kohlbergian perspective propose three conceptual levels of moral understanding: general schemas, intermediate concepts and codes (Rest, Narvaez, Bebeau, & Thoma, 1999; Thoma, 2006). Located within Rest's four component model (Rest, 1983), intermediate concepts are viewed as falling between the general "bedrock concepts" defined by Kohlberg's stage system or Defining Issues Test (hereafter the DIT) schema scores, and surface-level ethical

Correspondence should be addressed to Stephen Thoma, Educational Psychology, The University of Alabama, 310 Carmichael Hall, Tuscaloosa, AL 35406, USA. E-mail: sthoma@bamaed.ua.edu

codes that describe specific prescriptions and prohibitions (see Figure 1). For example, in the professions, intermediate concepts include topics such as informed consent, privacy, and professional autonomy. These concepts are typically the focus of professional education programmes and are described as central to the ethical life of a professional (Bebeau & Thoma, 1999).

Intermediate concepts (hereafter ICs) in the moral domain were first suggested as a partial response to critics of Kohlbergian-based professional ethics programmes. These critics argued that measures of Kohlberg's theory were unable to adequately assess ethics training due to the highly abstract properties of the stages they measure. For instance, Strike (1982), noted that individuals could have a good grasp of abstract moral schemas yet lack an adequate understanding of specific moral concepts that apply to day-to-day moral functioning. To these critics, moral schemas in the Kohlbergian tradition are necessary but not sufficient explanations of moral reasoning in context. Typically, intermediate concepts are the primary topics within professional ethics education programmes

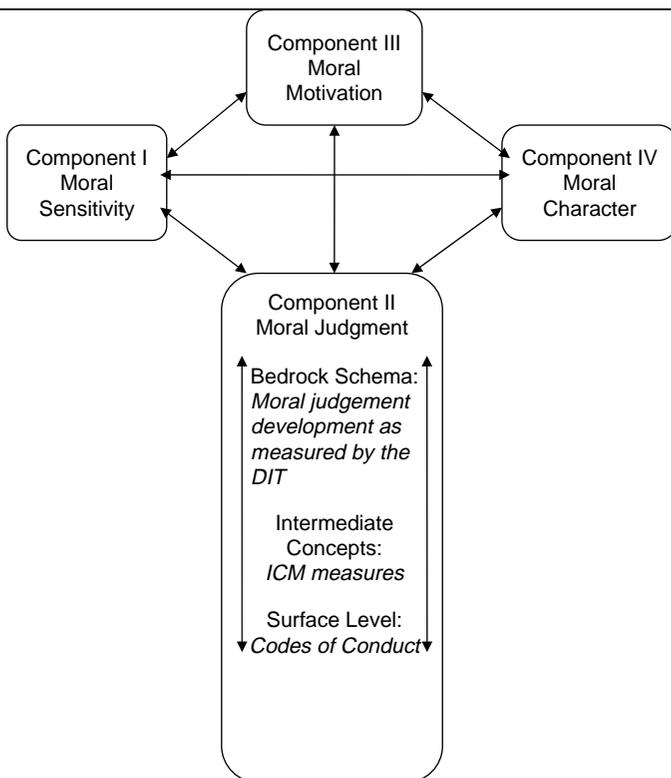


Figure 1. Intermediate concepts within Rest's four component model.

and have significant face validity in the profession. Similarly, at the level of professional ethics education, there is general consensus on what constitutes adequate and inadequate responses based on intermediate level concepts. Note that this consensus does not mean that the professionals tend to arrive at a single best application of an intermediate concept in any given situation. Indeed, professionals find it very difficult to arrive at a single best response or definition. However, there is surprisingly good agreement on the *set* of acceptable and unacceptable applications of the concept (e.g., Bebeau & Thoma, 1999). Preliminary work supports the utility of ICs as forming the basis for an assessment of discipline specific moral reasoning (e.g., Bebeau & Thoma, 1999; see also Turner, 2008).

Although measures of ICs have been found to provide unique insight into the moral judgement process and are particularly effective in assessing ethics education (e.g., Bebeau & Thoma, 1999), the empirical support for intermediate concept measures is currently limited to young adults in professional programmes. This limitation has been noted and some have questioned the claim that ICs define a generalized aspect of the moral reasoning process (e.g., Walker, 2002). To these critics, it is more prudent to view ICs as an artefact of the professional setting and associated well-established sets of moral considerations. Thus, one goal of the current paper is to test the theoretical claim that ICs can be generalized to other non-professional settings.

Taken together, the purpose of the article is three-fold. First, we describe the process of developing a generic IC measure designed for adolescent populations. This discussion will include the steps taken to generate dilemmas, items, and soliciting expert opinion. Second, we provide preliminary evidence for the reliability and validity of the measure by testing whether the summary indices relate to established measures of moral judgement development, are sensitive to age education trends across adolescence, and can distinguish students who have been placed in after school suspension for rules violations from comparable high school students. Third, in addition to testing whether the notion of ICs have a more generic application, we hope to provide an outcome measure that is appropriate for character education programmes and built on a strong theoretical and empirical foundation.

METHODS

Participants

Participants for this study were drawn from three populations. The first was a sample of 358 college students from a large south-eastern university who were asked to respond to the Adolescent Intermediate Concept Measure (hereafter AD-icm) as part of a larger study. All of these students were enrolled in an introductory human development course and received course credit for

participation. Given that the course is designed for entry-level freshman the majority of the sample were freshman and sophomores, however, juniors and seniors were also well-represented. Consistent with class demographics, the sample over selected women (69%) and social science majors. In addition, 21% of the students identified themselves as African Americans. The second sample ($n = 125$) was drawn from an urban high school in northern Georgia and a rural high school in Alabama. Participants were balanced by gender (52% female) and predominantly white (77%) In addition, 62% of the sample was in 9th and 10th grades. A third sample ($n = 15$) was collected in an Alabama high school district and represented students in after school suspension. These students had been placed in suspension for a variety of rule infractions but had in common a history of acting out in school. These students were overwhelmingly male (80%).

Measures

Defining Issues Test (DIT). The DIT (Rest, 1979) consists of six dilemmas; each followed by 12 items. Participants are first asked to take the role of the protagonist in the story and decide what he/she ought to do, and are then asked to rate and rank the items in terms of their importance in interpreting the moral dilemma. The traditional summary score of the DIT has been the “P” score, calculated from ranking data and attending to items keyed to Kohlberg’s Stages 5 and 6 (Kohlberg, 1969). A newer N2 score used in this study is an improvement over the P score as an overall estimate of moral judgement development (Thoma, 2006). More recently, the constructs measured by the DIT has been reinterpreted (Rest et al., 1999). Based upon large-sample analyses, it appears that the DIT measures three developmentally ordered schemas: personal interest (incorporating aspects of Kohlberg’s Stages 2 and 3), maintaining norms (closely aligned with Kohlberg’s Stage 4) and post-conventional schema (the traditional P score mentioned above). The validity and reliability of the DIT is fully discussed in Rest et al. (1999; see also Thoma & Bebeau, 2011).

Adolescent intermediate concepts measure (AD-icm). In order to transport the intermediate concepts measure 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 which represent plausible and implausible action choices and justifications.

Identifying the concepts. Following the tradition of using the curricular focus of educational programmes to help identify the most central intermediate concepts (e.g., Bebeau & Thoma, 1999) we focused on current character education programmes. Across programmes, the vast majority focus on the

virtues. For instance Lickona (1991) notes two core concepts that should reflect character: respect and responsibility. He further suggests that character incorporates concepts of honesty, fairness, tolerance, prudence, self-discipline, and courage. Comparable lists have been incorporated in other character education programmes (e.g., Arthur, 2008; Benninga, 1998, Bonner Foundation Conference papers).

We recognize that there is a historic tension between virtue-based and cognitive developmental models of moral education (e.g., Kohlberg, 1970) and it is not our intention in developing the AD-icm to contrast these two perspectives. In our view, the virtues as used in contemporary moral education are similar to established intermediate concepts in the professions (e.g., due process and informed consent), in that these concepts of character can be viewed as being understood by the individual based on his/her moral judgement processes interacting with contextual factors including training, experience, precedent, and cultural definitions. Thus, we suggest that there is a conceptual overlap between what are called intermediate concepts in the professional literature and the virtue-based concepts as used in the character education literature.

Identifying the target vignettes. Following the decision to frame the measure using the virtues as the intermediate level, the next step in the creation of the AD-icm was to identify the dilemmas used to highlight an application of each concept. A number of steps were employed to develop these dilemmas and in each step care was taken to solicit input from adolescents in the hope of maximizing the relevance of the resulting vignettes. First, 50 upper division high school students were asked to review a list of virtues and write real-life stories that highlight each concept. The results of this exercise ranged from highly creative and detailed stories to short and stereotypical responses. We then reviewed and sorted the stories by concept looking for common themes and situations. From these sorts, student responses were combined to create a set of stories that were relatively uniform in length and complexity. The resulting vignettes were then presented to 38 high school seniors and 36 college freshman who were asked to rate each story on realism and plausibility. Further, these students were asked to generate action choices for the story protagonist and supply justifications for these choices. We reviewed the plausibility data and discarded stories that were considered unrealistic. Following this process, we identified seven stories that represent one of the character concepts (see Table 1).

Developing the items. Having identified a set of stories, the next step in the measurement design phase was to develop a list of plausible action choices and justifications for each story. We began with the action choices and justifications identified during the dilemma development phase. These responses were sorted by type and a list of possible items was generated for each story. A small group

TABLE 1
The final list of dilemma themes

<i>Virtue</i>	<i>Short description</i>
Fairness	Whether or not to fire a friend who is the weakest worker
Responsibility	Whether or not to avoid a commitment and go out with friends
Loyalty	Supporting someone when it is difficult to do so
Self-discipline	Continuing to prepare for an important exam or go on a class trip
Honesty	What to do when friends cheat in school
Courage	Sticking to one's beliefs at the cost of recognition and success
Respect/loyalty	Fulfilling one's personal commitment to a mentor when abandoning him might result in greater personal gain

Note: The full AD-icm can be viewed at <http://www.ethicaldevelopment.ua.edu/adolescent-icm>.

($n = 20$) of college freshman reviewed the list of items and rated each proposed action choice and justification on a 5-point scale ranging from *Highly plausible* to *Highly implausible*. In addition, these students were asked to generate choices and justifications that they thought were absent from the lists. We then removed and/or altered items that were problematic and added the student nominations where appropriate. No attempt was made to standardize the number of choices or justifications for each story. Thus, some stories had fewer choices and justifications than others. Guiding this decision was the view that item realism was more important than item balance and 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 high school teachers, adolescents who have successfully manoeuvred 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 (e.g., Eccles, Lord, & Buchanan, 1996), we decided to emphasize social-science expertise over general experience and contact. Further, we felt that graduate students were not too removed from the cohort under study and were reasonably expert in their understanding of adolescent development.

As a first step, twenty graduate students were asked to rate each of the AD-icm items in terms of acceptability. Specifically, raters were asked to consider how they would judge an adolescent endorsing the various actions

(justifications)—would the action be acceptable, unacceptable or neutral? Items with good inter-rater agreement (75% raw agreement) were assigned the appropriate label. Items falling short of this agreement level were inspected and reworked as needed. A second sample of 24 students repeated the process. At the end of these two review cycles, all of the action choices and justifications for each story were reliably rated in one of the three categories (i.e., acceptable, unacceptable or neutral).

In a final step, we assessed the reading level of the items and stories. Consistent with the intended population the reading level for the measure was scored between the 4th and 5th grade (typically 9- and 10-year-olds) using the Flesch–Kincaid method.

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. Given that one of the purposes of this study was to assess the generalizability of the ICM approach, the adopted structure of the measure followed other IC approaches (e.g., Bebeau & Thoma 1999). Specifically, and after reading the story, we ask participants to rate a set of action choices on a 5-point scale (*I strongly believe that this is a good choice to I strongly believe that this is a bad choice*). After rating each action choice, participants then are asked to rank the three best choices and two worst choices. Following the action choice ranking task, participants rated the justification items on a similar 5-point scale (*I strongly believe that this is a good reason to I strongly believe that this is a bad reason*). Finally, the justification items were ranked using the same process as before (e.g., identify the three best and two worst). This process was then repeated for each of the seven stories resulting in the participant's assessment of the best and worst choices and justifications across stories.

Consistent with previous ICM measures the primary indexes of the measure focus on the ranking data. Ranks have been noted to provide superior trends presumably because the individual not only rates each item but then must consider the items as a set when identifying the best/worst items (e.g., Rest, 1979).

For the AD-icm higher scores reflect a ranking pattern in which the participants and experts agree. That is, if the participant selects the expert defined acceptable items as the best choices and justifications and in turn, identifies as worst choices and justifications the same way the experts rate the item, then he/she will receive a high score. By contrast, failing to match the experts reduces the scores. Following this procedure four intermediate scores are generated: Action choices—good and bad; justifications—good and bad using a weighted sum based on the ranking position. Specifically, first ranks are given one point more in the overall summary than later rank (i.e., for good choices and justifications: 1st rank = 3 points, 2nd rank = 2 and 3rd rank = 1, for 6 points possible). Across

stories this procedure yields a total possible “good” score of 42 (6 points by 7 stories) and a bad score of 21 (3 points by 7 stories).

The four intermediate scores defined above were combined to create three summary indices: Totalgood, Totalbad and TotalICM. Totalgood is defined as the average of the good choices (Good actions) and justifications (good justifications) represented as a percentage of a perfect score (i.e., a Totalgood score of 0.50 indicates that the participant identified 50% of the experts’ acceptable items as good choices). Totalbad is defined similarly (e.g., bad actions + bad justifications). Finally, the TotalICM score represents the average across all good and bad actions and justifications taking into account the range differences of the good and bad scales (i.e., 42 vs. 21 total points). It is also presented as a percentage. Therefore, a high TotalICM score indicates that the participant is appropriately identifying both acceptable and unacceptable items. The possible range for each score is -1 (a pattern of responses that are contrary to the expert key) through 0 to $+1$ (a pattern of responses in agreement with the expert key).

Procedures

The two high school samples completed the AD-icm and a demographic questionnaire in a regularly scheduled class period under teacher supervision. Due to the length of the AD-icm and 45 min completion time on average, it was not possible to gather any additional data. The college student sample completed the DIT and AD-icm in a group setting outside of class during a scheduled data-gathering session. Given the length of the full assessment battery and to avoid fatigue, two separate sessions were required to complete all of the measures. The DIT was completed in the first sitting and the AD-icm during the second session. The sessions were separated by two weeks.

RESULTS

Psychometric properties of the AD-icm

Internal consistency estimates were computed on the two summary scores (Totalgood and Totalbad) and the overall index (TotalICM). To compute these estimates we used the good choices, bad choices, and the story composite for each of the seven stories. In all cases, coefficient α values indicated good internal consistency and a reasonably stable estimate of the various scales (Totalgood: $\alpha = .75$; Totalbad: $\alpha = .77$; TotalICM: $\alpha = .85$). Further, correlations between the intermediate scale scores were all moderate and in a positive direction (see Table 2). Thus, each scale offers a non-redundant source of information. These findings support treating the three scores as independent indexes of individual’s intermediate concept usage.

TABLE 2
Pearson product moment correlations between the AD-icm raw scales

<i>Scales</i>	<i>Act good</i>	<i>Act bad</i>	<i>Just bad</i>	<i>Just good</i>
Act good	1.00			
Act bad	.62	1.00		
Just bad	.58	.54	1.00	
Just good	.65	.53	.64	1.00

Notes: Act good = selecting appropriate action choices. Act bad = selecting inappropriate action choices. Just bad = selecting appropriate justifications. Just good = selecting inappropriate justifications. All correlations are statistically significant.

Construct validity

Pearson product-moment correlations between AD-icm and DIT summary scores were computed using the college sample. Consistent with the claim that both measures assess the moral domain, all correlations with the summary score (N2) were statistically significant, $r(310) = .30, .27, .29$ for TotalICM, Totalgood and Totalbad, respectively. It is interesting to note that correlations between the personal interest schema score and AD-icm summary scores were uniformly negative, $r(310) = -.27, -.22, -.29$ for TotalICM, Totalgood and Totalbad, respectively, suggesting that AD-icm scores increase as personal-interest schema scores decline. By contrast, correlations between the maintaining norms schema scores and the AD-icm scores were positive and statistically significant, $r(310) = .19, .19, .15$ for TotalICM, Totalgood and Totalbad, respectively. These values are similar to those reported for other ICM approaches (e.g., Bebeau & Thoma, 1999).

To further explore the relationship between the DIT and AD-icm summary scores, we blocked participants by their modal moral schema score and compared the resulting three groups on the AD-icm summary scores. One-way analyses of variance (ANOVAs) with post hoc comparisons indicated statistically significant overall effects, $F(2, 309) = 19.10, p < .05, \eta^2 = .11$; $F(2, 309) = 9.30, p < .05, \eta^2 = .06$; $F(2, 309) = 15.07, p < .05, \eta^2 = .08$ for Totalbad, Totalgood, and TotalICM, respectively. Consistent with the correlational findings, post hoc comparisons using Tukey's procedure showed that the group emphasizing personal interest schema obtained the lower ICM scores.

Age education trends on the AD-icm

An important question addressed by this study is whether age-educational groups are different on the AD-icm. Following the findings described above that indicate the AD-icm is related to DIT scores, we expected that there would be differences between groups throughout the high school years. Further, and consistent with the

view that the AD-icm is sensitive to groups who are different in acting out behaviours, we further expected that the in-school suspension group should have lower than average scores.

Table 3 presents the means and standard deviations for the seven groups on the major AD-icm scores. Group differences were observed on all three summary indexes, $F(6, 440) = 14.77, p < .01, \eta^2 = .20$, $F(6, 440) = 12.88, p < .01, \eta^2 = .18$, $F(6, 440) = 13.47, p < .01, \eta^2 = .19$ for TotalICM, Totalgood, Totalbad, respectively. Finding significant differences on each index, planned contrasts were used to test the comparisons of interest. These contrasts indicated that high school students scored lower than college students on all indexes and effect sizes were moderate to large, $t(440) = 7.46, p < .01, d = 0.71$; $t(440) = 7.24, p < .01, d = .69$; $t(440) = 6.64, p < .01, d = 0.63$ for TotalICM, Totalgood, and Totalbad, respectively. Younger high school students scored lower than upper division high school students on Totalbad, $t(440) = 2.20, p < .01, d = 0.21$; and reached a statistical tendency on TotalICM, $t(440) = 1.62, .10 > p > .05, d = 0.15$, but were not statistically different on Totalgood, $t(440) = 1.20, p > .05, d = 0.11$. No statistically significant differences were observed across the college groups.

Students who were suspended from regular classes scored lower than their high school peers on TotalICM, $t(440) = 2.12, p < .05, d = 0.20$, and Totalbad, $t(440) = 2.55, p < .01, d = 0.24$. Differences on Totalgood indicated a statistical tendency, $t(440) = 1.62, .10 > p > .05, d = 0.15$. Taken together, the ANOVA results support the view that the AD-icm scores differ by group. Further, the planned contrast analyses indicate that, on the whole, the hypothesized group differences were statistically confirmed with one exception: college students did not differ by age and year in college.

Observing that Totalbad scores were consistently lower than Totalgood scores and noting that Bebeau and Thoma (1999) found a similar trend in dentistry

TABLE 3
Group differences on AD-icm scales

<i>Group</i>	<i>N</i>	<i>Totalgood</i> <i>M (SD)</i>	<i>Totalbad</i> <i>M (SD)</i>	<i>TotalICM</i> <i>M (SD)</i>
School suspension	15	0.42 (0.19)	0.19 (0.27)	0.34 (0.20)
Lower division HS	74	0.48 (0.22)	0.31 (0.28)	0.42 (0.23)
Upper division HS	46	0.52 (0.23)	0.40 (0.31)	0.48 (0.25)
College freshman	80	0.63 (0.19)	0.53 (0.21)	0.59 (0.18)
College sophomores	131	0.65 (0.15)	0.53 (0.22)	0.61 (0.16)
College juniors	56	0.64 (0.14)	0.54 (0.19)	0.61 (0.14)
College seniors	45	0.65 (0.15)	0.53 (0.18)	0.60 (0.15)

Notes: Totalgood = Summary of good choices and justifications. Totalbad = Summary of bad choices and justifications. TotalICM = Summary of good and bad choices and justifications.

students, the difference between the two scores was assessed. A 2×7 repeated-measures ANOVA with Totalbad and Totalgood as the within-subjects factor and Group the between-subjects factor indicated that Totalbad scores were statistically lower than Totalgood scores, $F(1, 440) = 222.94$, $p < .01$, $\eta^2 = .34$ and $F(6, 440) = 14.93$, $p < .01$, $\eta^2 = .17$ for the item type and group effect, respectively (note that the range differences—i.e., 42–21 total points—on these measures cannot account for the obtained mean differences and, thus, the difference between averages is not due to a methodological artefact). In addition, a small but significant group \times item interaction was observed, $F(6, 440) = 3.03$, $p < .01$, $\eta^2 = .04$. Inspection of the means indicates that the interaction is due to a greater difference between scores in the younger groups.

Individual differences on the AD-icm

In addition to the group trends, additional analyses were conducted in order to assess whether gender or race are related to AD-icm scores. Given the demographics of the sample, race was a dichotomous variable defined by White and African-American students. Results of the analyses with race indicated no statistically significant differences either independently or in combination with age/educational groupings. By contrast, statistically significant moderate to large gender differences favouring females were observed across the three AD-icm scores ($d = 0.65, 0.66, 0.55$ for TotalICM, Totalgood and Totalbad scores, respectively). No interactions between gender and age/educational groupings were observed.

DISCUSSION

Results of this study indicate that the newly created AD-icm is sensitive to age educational groups across the high-school years. In addition, AD-icm scores are related to DIT scores in college populations providing some support for the claim that the constructs measured by both measures overlap. Coupled with the respectable psychometric properties of the measure, these findings offer preliminary support for the AD-icm as a viable measure of moral thinking in adolescent populations.

Whether or not the measure should be used in young adult populations is less clear given the absence of age trends in the college sample. However, we would note that the college sample was drawn from a freshman-level course and the presence of juniors and seniors may represent an atypical subset of upper division students. For instance, this group may be overly represented by students having more difficulty attaching to a course of study, or who were simply not following the proscribed plan of study (e.g., Pascarella & Terenzini, 2005). Further work with more representative college samples appears necessary before establishing an upper age limit for the AL-icm.

Finding that most of the movement on the measure was associated with the high-school years and noting the pattern of relationships with the DIT, one plausible interpretation of the AD-icm is that we have created a measure that is sensitive to the transition from personal interest to conventional thinking. This interpretation is not surprising given the reliance on adolescent informants in determining the different stories and items. Having a measure that captures this transition to conventional thinking is developmentally appropriate for adolescents and is consistent with the aims of many character education programmes in which the importance of conventions in framing moral thinking and character are explicit. As descendants of Kohlberg, we are mindful of the shortcomings associated with programmes of this type; however we also note the importance of conventional reasoning as a springboard for later development in social-moral thinking. Thus we do not view the measure as in some way limiting or overly narrow.

It is also important to note that the success of the AD-icm directly addresses the question of whether ICs are a generalized phenomenon or simply an artefact of the professional setting (e.g., Walker, 2002). The findings of this study offer support for the claim that intermediate concepts can be identified in more generic and non-professional settings. Additionally the ease with which the measurement system developed in Bebeau and Thoma (1999) was transported to adolescent issues and populations suggests the utility in using these procedures in further attempts to develop an IC measure. Having an established methodology should encourage the development of intermediate concepts in other settings and populations and further address the need for more contextualized measures of moral thinking.

Interestingly, the ability to identify bad choices and justifications appears more difficult for participants in this study. In addition, group differences on these variables were more pronounced than on the corresponding variables addressing the identification of acceptable items. As mentioned previously, the same pattern was observed using the dentistry intermediate concepts measure with young adults in the last year of dentistry training (Bebeau & Thoma, 1999). 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 have less formal experience deducing bad choices and justification. Whether this finding is a reflection of how our culture socializes its children or due to a more general developmental process, a continued focus on the difference between identifying good and bad choices seems especially warranted.

It was somewhat surprising to find a fairly large gender difference on the AD-icm. Typically, gender differences on objective measures of moral judgement development are small (e.g., Thoma, 1986; Walker, 2006). It may be women have an advantage because the stories developed for the AD-icm were more influenced by their input. Although we have no data to support this claim, our

recollection is that women responded to our early requests for stories and items with more detailed responses. To counter this possibility we were sensitive to gender in the reactions to the various stories during the development phase of the measure. However, it still may be that the dilemmas eventually selected are more familiar to women. Future research using the AD-icm should monitor and attend to the possibility of gender differences.

Finally, we note that over the last two decades many practitioners and politicians have raised concerns about the behaviour and character of our youth (e.g., Arthur, 2008; Benninga, 1998). Typically the means suggested for addressing these concerns focus on the schools and the need to add character education to the curriculum. Unfortunately, most of the suggested programmes are not built on a firm theoretical foundation and few appropriate measures are available for programme evaluation purposes (e.g., Bebeau, Rest, & Narvaez, 1999). In our view, the AD-icm overcomes many of the criticisms Bebeau and her colleagues identify: The measure is theory driven and represents moral thinking within specific contexts particularly salient to adolescents. Additionally, and extrapolating from studies using ICMs developed in professional populations (e.g., Bebeau & Thoma, 1999), the AD-icm may be more sensitive than traditional measures to the quality of everyday moral thinking and outcomes of ethics interventions. As such, the AD-icm represents a promising outcome measure for character education programmes.

REFERENCES

- Arthur, J. (2008). Traditional approaches to character education in Britain and America. In L. P. Nucci & D. Narvaez (Eds.), *Handbook of moral and character education* (pp. 80–98). New York, NY: Routledge.
- Bebeau, M., Rest, J., & Narvaez, D. (1999). Beyond the promise: A perspective of research in moral education. *Educational Researcher*, 28, 18–26.
- Bebeau, M., & Thoma, S. (1999). Intermediate concepts and the connection to moral education. *Educational Psychology Review*, 11, 343–360.
- Benninga, J. (1998, May 22–23). *Bonner Center for Character Education and Citizenship Assessment Forum Final Report*. Fresno, CA: California State University.
- Eccles, J. S., Lord, S., & Buchanan, M. (1996). School transitions in early adolescence: What are we doing to our young people? In J. A. Graber, J. Brooks-Gunn & A. Peterson (Eds.), *Transitions through adolescence* (pp. 251–284). Mahwah, NJ: Erlbaum.
- Kohlberg, L. (1969). Stage and sequence: The cognitive developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory* (pp. 347–480). Chicago, IL: Rand McNally.
- Kohlberg, L. (1970). Education for justice: A modern statement of the Platonic views. In N. F. Sizer & T. R. Sizer (Eds.), *Moral education: Five lectures* (pp. 57–83). Cambridge, MA: Harvard University Press.
- Lickona, T. (1991). *Educating for character*. New York, NY: Bantam.
- Pascarella, E., & Terenzini, P. (2005). *How college affects students: A third decade of research*. San Francisco, CA: Jossey-Bass.

- Rest, J. R. (1979). *Development in judging moral issues*. Minneapolis, MN: University of Minnesota Press.
- Rest, J. R. (1983). Morality. In P. H. Mussen (Series Ed.), J. Flavell & E. Markman (Vol. Eds.), *Handbook of child psychology. Vol. 3, Cognitive development* (4th ed., pp. 556–629). New York, NY: Wiley.
- Rest, J. R., Narvaez, D., Bebeau, M., & Thoma, S. J. (1999). *Post conventional moral thinking: A neo-Kohlbergian approach*. Mahwah, NJ: Erlbaum.
- Strike, K. A. (1982). *Educational policy and the just society*. Chicago, IL: University of Illinois Press.
- Thoma, S. J. (1986). Estimating gender differences in the comprehension and preference of moral issues. *Developmental Review*, 6, (pp. 165–180). (Reprinted in Jacklin, C.N. (1992). *The psychology of gender*. London, UK: Edward Elgar Publishing).
- Thoma, S. J. (2006). Research using the Defining Issues Test. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (pp. 67–92). Mahwah, NJ: Erlbaum.
- Thoma, S. J., & Bebeau, M. (2011). *Guide to the DIT-1*. Available from the Center for the Study of Ethical Development, The University of Alabama, Tuscaloosa, AL.
- Turner, M. (2008). *The development and testing of the Army Leader Intermediate Ethical Concept Measure*. Unpublished dissertation, The University of Alabama, Tuscaloosa, AL.
- Walker, L. (2002). The model and the measure: An appraisal of the Minnesota approach to moral development. *Journal of Moral Education*, 31, 353–367.
- Walker, L. J. (2006). Gender and morality. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (pp. 93–118). Mahwah, NJ: Erlbaum.