

Does the Defining Issues Test Measure Psychological Phenomena Distinct From Verbal Ability?: An Examination of Lykken's Query

Cheryl E. Sanders, David Lubinski, and Camilla Persson Benbow
Iowa State University

This study examined the incremental validity of the Defining Issues Test (DIT), a test purporting to measure moral reasoning ability relative to verbal ability and other major markers of the construct of general intelligence (*g*). Across 2 independent studies of intellectually precocious adolescents (top 0.5%), results obtained with the DIT revealed that gifted individuals earned significantly higher moral reasoning scores than did their average-ability peers; they also scored higher than college freshmen, who were 4 to 5 years older. The relative standing of the intellectually gifted adolescents on moral reasoning, however, appears to be due to their superior level of verbal ability as opposed to any of a number of the other psychological variables examined here. The hypothesis that the DIT is conceptually distinct from conventional measures of verbal ability was not confirmed. Investigators conducting subsequent studies involving the assessment of moral reasoning are advised to incorporate measures of verbal ability into their designs, thereby enabling them to ascertain whether moral reasoning measures are indeed capturing systematic sources of individual differences distinct from verbal ability. This idea also is relevant to other concepts and measures purporting to assess optimal forms of human functioning more generally (e.g., creativity, ego development, and self-actualization).

In the social sciences, measures do not always assess what they purport to measure, and the causal determinants of our most favorite constructs and outcomes do not always fit prior expectations. All too frequently in social science research, theoretically appealing constructs are assessed and studied without attending to competing variables that might be causally linked to their status as well as the criteria they predict. Socioeconomic status (SES), for example, is a variable that social scientists frequently assume to be causally related to a host of psychologically important phenomena (see the writings of Humphreys, 1991, and Meehl, 1970, 1971a, 1971b, on this topic). In school and work settings, for example, SES is presumed to exert a profound causal role in determining the outcomes on conventional achievement criteria (cf. Humphreys, 1991). More powerful variables exist, however, for predicting academic and vocational criteria (e.g., human abilities; Humphreys, 1992; Lubinski & Dawis, 1992; Schmidt, 1994). Yet, they are seldom studied concomitantly with SES (Humphreys, 1991). Indeed, the practice

of inferring causal links between SES—the “master variable” of sociological inquiry (Gordon, 1987)—and its many correlates, without evaluating competing correlated factors (such as ability), has been referred to as the *sociologist's fallacy* (Jensen, 1973). This common methodological shortcoming is an example of underdetermined causal modeling—and the social sciences are dotted with several others.

One of the more striking examples of this in contemporary psychological research is the tendency for researchers to evaluate the importance of students' self-efficacy or its manipulation for choosing to embark on conceptually demanding educational or vocational paths, without simultaneously assessing relevant ability requirements necessary for performing competently in the targeted disciplines (for a number of examples, see Betz & Fitzgerald, 1993). Another example can be found in research on parent perceptions of their sons' and daughters' strengths and weaknesses. Gender differentiating expectations are frequently interpreted as a function of sex role stereotyping (e.g., Jacobs & Eccles, 1992); although this may be so, more definitive conclusions would be obtained if *objective* measures of the rated skill domain under analysis were concomitantly assessed (say, for example, *ratees'* mathematical reasoning ability), on which parents' perceptions and ratings are at least partly based. We would then be in a position to determine whether genuine gender differentiating competencies were actually observed and rated with precision, or whether parents' ratings were indeed moderated by sex role stereotypes and, as such, are systematically biased.

Erroneous suppositions regarding presumed causal paths are

Cheryl E. Sanders, David Lubinski, and Camilla Persson Benbow, Department of Psychology, Iowa State University.

This article was based on a dissertation submitted by Cheryl E. Sanders to Iowa State University in partial fulfillment of the Doctor of Philosophy degree. An earlier version of this article profited from comments and suggestions from Lloyd G. Humphreys, Paul E. Meehl, and Julian C. Stanley.

Correspondence concerning this article should be addressed to David Lubinski, Department of Psychology, W112 Lagomarcino Hall, Iowa State University, Ames, Iowa 50011-3180.

clearly found in discussions of well-known environmental measures. The Moos and Moos (1986) Family Environment Scale (FES) has almost always been interpreted as indexing an *exogenous* source of environmental causality. Robert Plomin and his colleagues have revealed, however, that the FES manifests significant heritabilities across twin and adoption studies (Plomin & Bergman, 1991; Plomin, Reiss, Hetherington, & Howe, 1994) and, hence, an *endogenous* source of genetic variation. As a result, investigators conducting research with the FES and similar measures must now modify their causal conjectures accordingly.

Such problems of unevaluated competing explanations are ubiquitous not only in psychological research (cf. Scarr, 1992; Scarr & McCartney, 1983), but also at more fundamental levels of measurement. This happens most conspicuously when investigators attempt to establish the construct validity of new, innovative instruments (Dawis, 1992; Lubinski & Dawis, 1992).¹ Measures of favorite constructs (e.g., creativity, ego development, moral reasoning) are frequently constructed and "validated" within elaborate networks of criterion variables and experimental manipulations, without ever considering the possibility that other *existing measures* might account for the same correlational and experimental findings as well as, or perhaps more comprehensively than, the investigator's purported ("master") construct. Just as predictor-criterion correlations can have competing causal interpretations, innovative measures of constructs may actually reduce to weak measures of other constructs better assessed with preexisting instruments. In what follows, we examine an instance of this possibility.

Assessing Moral Reasoning

In an eye-opening methodological treatment of a number of problems associated with psychological research, such as those illustrated above, David Lykken (1991, p. 35) posed the following query: "One can reasonably wonder whether many of the interesting findings obtained in research on Kohlberg's (1984) Stages of Moral Development would remain if verbal intelligence had been partialled out in each case."

The purpose of the present study was to examine Lykken's (1991) query. We evaluated the Defining Issues Test (DIT), an objective measure of moral reasoning ability that is based on Kohlberg's system (Rest, 1979a, 1979b), in the context of a number of conventional ability measures. Our analysis was aimed at answering two questions (one is methodological, the other is substantive). Both questions are addressed concurrently: (a) how is the uniqueness of the DIT best established? and (b) does the psychological importance of the DIT reduce to its overlap with verbal ability? The latter question, Lykken's query, is actually not without some empirical support.

A number of studies have reported substantively meaningful correlations between the DIT and intellectual abilities. Rest (1979a, 1979b), for example, reported correlations between the DIT and intelligence in the .20-.50 range. Also, in earlier empirical work, Kohlberg (1969) himself actually reported correlations between his Moral Judgment Interview and general intelligence, ranging from .30 to .50. Furthermore, discussions of the association between moral reasoning and intelligence have appeared for decades in the psychological literature (e.g.,

Abel, 1941; Boehm, 1967; Durkin, 1959; Kohlberg, 1969; Perry & Krebs, 1980; Piaget, 1932; Simmons & Zumpf, 1986; Terman, 1925; Whiteman & Kosier, 1964). Yet none have specifically examined the uniqueness of moral reasoning measures relative to conventional markers of intelligence in the context of meaningful psychological criteria (Colby & Kohlberg, 1987). This is precisely what we did.²

We used a host of psychological variables to predict DIT scores *after* partialing out markers of general intelligence—verbal and quantitative abilities and a nonverbal measure of general intelligence. A variety of markers of general intelligence was used, inasmuch as Rest (1979a, 1979b) has reported that math and science test scores seem to predict DIT scores as well as language, vocabulary, and social science test scores. So a corollary hypothesis that emerges is that it is the communality running through heterogeneous collections of cognitive tests, as opposed to the specificity of more circumscribed markers of general intelligence, that overlaps with the DIT.³ Is it possible then, that, beyond its overlap with verbal ability (or more global measures of general intelligence), the variance shared between the DIT and relevant psychological criteria is nugatory?

¹ Meehl (1990a, 1990b), among others, has commented on this problem.

² Although there are other measures of moral reasoning available, the DIT appears to be an especially good choice for evaluating Lykken's query. Indeed, in a recent evaluation of this instrument, Rest and Narvaez (1994) concluded: "The fact that the DIT is one of the easiest tests to administer and score (being multiple-choice and computer-scored) should not be held against it. Despite its ease of use, there is no other program of research with other instruments that has produced clearer findings or more useful information about professional ethics. Although other instruments usually involve more pain, there is not inevitably more gain." (p. 214)

³ Our hypothesis that the meaningful individual differences assessed by the DIT reduce to the dominant dimension defined by the communality of many different kinds of cognitive tests, namely general intelligence (Humphreys, 1979), is also reinforced by auxiliary data. The construct of general intelligence actually does surface as predictively central in many contexts outside of educational and vocational settings and across more behavioral ecologies than many psychologists realize (Brand, 1987; Humphreys, Rich, & Davey, 1985; Lubinski & Dawis, 1992). As the late Starke R. Hathaway use to tell all of his clinical advisees, "We tend to treat general intelligence as if it only operated in educational and vocational contexts; but actually, it saturates everything we do and is a salient aspect of personality" (Paul E. Meehl, July 1993, personal communication). Readers also might find of interest the remarkable consensus among leading psychometricians on how human intellectual abilities are organized (cf. Ackerman, 1987, 1988; Carroll, 1993; Gustafsson, 1988; Humphreys, 1979; Snow & Lohman, 1989; Vernon, 1961). The general consensus is that the intellectual repertoire is organized hierarchically, with the construct of general intelligence at the vertex and accounting for approximately half of the variance running through heterogeneous collections of ability tests in the full range of talent. This dimension of individual differences defines the level of sophistication of the intellectual repertoire; but the specificity of its three major markers (viz., verbal, quantitative, and spatial group factors) contain incremental validity beyond the general factor—and we actually demonstrate this for verbal ability in the present study. See Humphreys, Lubinski, and Yao (1993) and Lubinski and Dawis (1992) for further and more detailed discussions on the nature and organization of human intellectual abilities and their many correlates.

In the present study, in addition to a variety of cognitive measures, criterion variables were drawn from a number of other psychological domains (reported to be significantly related to the DIT in other research): leisure activities (Biggs & Barnett, 1981; Duffy, 1982; Laubscher, 1988), personality (Cauble, 1976; Hanson & Mullis, 1985), values (Lockley, 1976), and a variety of background and family characteristics (Parikh, 1980; Rest, Cooper, Coder, Masanz, & Anderson, 1974; Snarey, Reimer, & Kohlberg, 1984). We wanted to determine the nature and strength of the relationship between these various measures and the DIT, after removing the variance shared by markers of general intelligence through partial correlation.

Method

Two studies were included in this research for purposes of replication. This cuts down the possibility of interpreting chance correlations and partial correlations. Both studies focused on gifted youths who attended the CY-TAG (Challenges for Youth—Talented and Gifted) and Iowa Governor's Summer Institute (IGSI) programs at Iowa State University. These are summer programs designed to better meet the educational needs of intellectually gifted adolescents (Benbow, 1990; Benbow & Stanley, 1983; Lubinski & Benbow, 1994). Study 1 was conducted in 1990 and included 147 males and 121 females; Study 2 was carried out in 1991 with 136 males and 119 females. Study 2 followed Lykken's (1968) three-tiered nomenclature for conducting replications in psychological research and constituted a literal replication of Study 1.

Individuals are eligible for CY-TAG and IGSI programs if they are currently enrolled in 7th to 10th grades and if they qualify as intellectually gifted. Students are selected for CY-TAG on the basis of College Board Scholastic Aptitude Test (SAT) scores or scores on the American College Test (ACT) assessment, tests normally administered to 11th and 12th grade students intending to enroll in college. Additional requirements for CY-TAG involve earning one of the following test scores as a 7th grader: ≥ 500 on the SAT-Math (SAT-M) subtest, ≥ 430 on the SAT-Verbal (SAT-V) subtest, ≥ 930 on SAT-M + SAT-V, or ≥ 20 on any ACT subtest. Minimum SAT and ACT scores earned by CY-TAG participants at age 12 to 13 are comparable to the average score received by college-bound high school senior males. Although selection for the IGSI was not based on SAT or ACT scores, many such students had taken these tests. Those who had earned scores comparable to CY-TAG participants were included in the present research. Thus, the sample represents approximately the top 0.5% in intellectual ability as measured by the SAT or ACT. The gifted youths volunteered for the research in return for a subsidy to the program. Because not all students who qualified for CY-TAG were administered the SAT, our analyses with SAT-M and SAT-V scores consist of only 92 males and 72 females (Study 1) and 102 males and 83 females (Study 2).⁴

To establish baselines for evaluating the overall DIT performance of our intellectually gifted adolescents, we include means and standard deviations from the DIT manual as well as from two separate studies conducted at Iowa State University for other purposes. One of these latter two groups consisted of individuals of equivalent chronological age to the gifted youths (30 male and 27 female 12- to 14-year-olds) but of average ability (according to their Iowa Test of Basic Skills scores). Their SES was not significantly different from our gifted participants. These students were paid \$5.00 each for their participation. The second group, consisting of 49 male and 83 female college freshmen, served as an additional comparison. They received extra credit in an introductory psychology class for their participation.

Instrumentation

Materials used were selected items and scales from the following instruments: the DIT (Rest, 1979a, 1979b), FES (Moos & Moos, 1986),

Adjective Checklist (ACL; Gough & Heilbrun, 1983), Study of Values (SOV; Allport, Vernon, & Lindzey, 1970), Raven's Progressive Matrices (Raven, Court, & Raven, 1977), Background Questionnaire for CY-TAG Students, and an activities questionnaire. Descriptions of the variables used from each instrument are provided next.

DIT. We assessed moral reasoning with the DIT, a standardized instrument based on Kohlberg's (1984) theory of moral development and constructed by Rest (1979a, 1979b). It is an objective instrument consisting of six story dilemmas, each describing a situation requiring an ethical decision. Associated with each dilemma are 12 statements representing a particular stage of moral judgment. The participants are asked to rate the importance of each statement and to select the four most important issues, ranking them in order of importance. Scores are based on the relative importance participants place on stage-related statements (cf. Rest, 1973, 1975, 1976).

FES. We assessed the social-environmental characteristics of family with the FES (Moos & Moos, 1986). The FES consists of 10 scales, which are classified into three domains: relationships, personal growth, and system maintenance. The relationship class is made up of the Cohesion, Expressiveness, and Conflict subscales. This cluster assesses the extent to which family members are supportive, open, and expressive with each other. The personal growth cluster includes Independence, Achievement Orientation, Intellectual-Cultural Orientation, Active-Recreational Orientation, and Moral-Religious Emphasis subscales. This group of scales focuses on the degree to which family members are assertive, self-sufficient, and interested in political, social, intellectual, religious, cultural, and recreational activities. The system maintenance cluster includes Organization and Control subscales. This class involves how important structure and organization are in the family unit.

There are three forms of the FES: the Real, Ideal, and an Expectations form. We used the Real form in this research. It measures the students' perceptions of their family environment as it currently exists.

ACL. We used the ACL (Gough & Heilbrun, 1983) to assess personality attributes. The ACL is composed of 300 adjectives used to form 37 scales, which, in turn, are categorized into five classes. The first class, measuring needs, consists of: achievement, dominance, endurance, order, intraception, nurturance, affiliation, heterosexuality, exhibition, autonomy, aggression, change, succorance, abasement, and deference. Topical scales include: counseling readiness, self-control, self-confidence, personal adjustment, ideal self, creative personality, military leadership, masculine attributes, and feminine attributes. Transactional analysis scales consist of: critical parent, nurturing parent, adult, free child, and adapted child scales. Finally, the origence-intellectence scales, which assess one's tendency to reason abstractly as well as creatively, include: high origence, low intellectence; high origence, high intellectence; low origence, low intellectence; and low origence, high intellectence.

SOV. We used the SOV (Allport et al., 1970) to assess six basic aspects of personality in an ipsative fashion. The six values include theoretical, economic, aesthetic, social, political, and religious values. The SOV is based on the view that people's personalities are assessed by investigating their value systems. Although the SOV is a self-administered test designed primarily for college students or adults who have had some college or equivalent education, use of the instrument with participants of this research was deemed acceptable because of their high ability and the long tradition of using this instrument with such students (cf. Keating, 1974; Lubinski & Benbow, 1992; Lubinski, Benbow, & Sanders, 1993).

Raven's Advanced Progressive Matrices. We used the Raven's matrices (Raven, Court, & Raven, 1977) as a test of nonverbal reasoning

⁴ Mean levels on the DIT for these participants were actually all a bit higher than those reported in Table 1, but not significantly so.

ability. It consists of 36 items. Each item involves a pattern of figures and options for solving a relational problem.

Background questionnaire for CY-TAG students. The background questionnaire for CY-TAG students is a general information survey completed by all participants of CY-TAG and Iowa Governor's Institute programs. Demographic information, as well as questions pertaining to students' feelings and opinions, are included in the questionnaire. The following four items were used from this questionnaire as indexes of SES: paternal educational level, maternal educational level, paternal occupation, and maternal occupation.

Activities questionnaire. We used an activities questionnaire to assess extent of participation in various leisure activities and hobbies. Factor analytic investigations of this instrument have revealed five interpretable dimensions (all of which will be used in the present investigation): Involvement in Nonfiction Reading, School Clubs, Math/Science-Related Activities, Video Games, and Fiction Reading.

Data collection. The questionnaires were administered by mail, before students actually attended the summer programs, whereas the DIT and Raven's matrices were administered in classroom settings once the students arrived at CY-TAG or IGSI.

Results

For efficiency in exposition, results obtained from Study 1 and Study 2 are presented concurrently. Means and standard deviations of the DIT are reported in Table 1, by gender, along with data collected on average-ability, age-equivalent peers and college students (who were 4 to 5 years older than the other two groups). In both studies, the gifted adolescents appeared to be more advanced in terms of their moral reasoning abilities (and the other abilities) than both their age-equivalent peers and the college sample.

Next, for both studies, we correlated all 62 criterion measures with the DIT, by gender. As one would imagine from Type I error expectations and the number of criterion measures used, we obtained a number of statistically significant correlations in both studies. Yet only a small subset of the statistically significant correlations found in Study 1 were replicated in Study 2, for both genders. This underscores the importance of conducting literal replications in studies of this kind. Table 2 consists of the intercorrelations of all the statistically significant correlations of the DIT found in both Study 1 and Study 2 across both genders.

As can be noted, correlations between the cognitive measures and the DIT are the only ones that held up across both studies. Moreover, the DIT appears to be as highly correlated with the measures of intellectual ability as these ability measures are with each other. Although these correlations are modest, we must consider that our sample consisted of a highly restricted range of talent. Thus, the magnitude of these correlations is attenuated appreciably. This is underscored by the light correlations between SAT-V and SAT-M. In college-bound high school seniors (a much less restricted range of talent, but nevertheless restricted) these two scales correlate in the .60-.70 range, compared with the .24-.61 range observed in this study. This speaks to the high degree of overlap among content-distinct intellectual measures, which is frequently underappreciated when working with highly select samples (Benbow, 1988, 1992; Lubinski & Dawis, 1992). This is also why we chose to employ a number of content-distinct markers of general intelligence—to illustrate how the same source of common variance (frequently denoted as *g*) can manifest itself in different ways (Lubinski & Dawis, 1992).

Following the above analysis, we combined the genders in both Study 1 and Study 2 to form two gender-mixed groups. We then correlated the DIT with all 62 criterion measures for these two samples. In the gender-combined samples, three noncognitive criterion variables manifested replicable correlations with the DIT across both studies (with Study 1, followed by Study 2 correlations given in parentheses): intellectual-cultural orientation (.19, .12), creative personality (.11, .15), and video game playing (-.17, -.18). The correlations for the three cognitive variables and the DIT in the gender-combined sample were: SAT-V (.30, .45), SAT-M (.27, .25), and Raven (.19, .20).

To evaluate whether the replicated covariation between the DIT and the noncognitive variables was distinct from that of our markers of general intelligence, we conducted 18 stepwise multiple regression analyses (following a forward selection procedure). We ran all possible two-predictor pairs between the noncognitive measures (intellectual-cultural orientation, creative personality, and video games) and each of the three markers of general intelligence (SAT-V, SAT-M, and Raven's matrices), across both studies. Given that the cognitive mea-

Table 1
Means and Standard Deviations of DIT-P% Scores of Gifted, Junior High, Average Ability, and College Students

Sample	Group			Males			Females		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Study 1 gifted youths	268	34.9	12.9	147	33.3	12.4	121	36.7	13.3
Study 2 gifted youths	255	35.6	12.4	136	34.4	12.0	119	36.9	12.8
Average ability 12- to 14-year-olds	57	22.4	10.6	30	22.5	10.2	27	22.1	11.1
College freshmen ^a	131	30.0	12.7	49	30.1	13.0	83	30.0	12.5
Junior high norms ^b	1,322	21.9	8.5						
Senior high norms ^b	581	31.8	13.5						

Note. DIT = Defining Issues Test. DIT-P% = most commonly employed index of moral reasoning assessed by the DIT.

^a Iowa State University freshmen.

^b Data from the DIT manual (Rest, 1979b).

Table 2
*Intercorrelations Between Cognitive Variables
 and DIT-P% Scores*

Variable	1	2	3	4
1. SAT-V	—	.24	.05	.25
2. SAT-M	.44	—	.43	.39
3. Raven	.43	.55	—	.26
4. P% score	.28	.43	.15	—
	.56	.29	.19	

Note. Correlations for females are above the diagonal; those for males are below. The top correlation in each row is from Study 1 (1990); the bottom correlation is from Study 2 (1991). All P% score correlations are statistically significant at $p < .05$ or beyond. Sample sizes for the Scholastic Aptitude Test (SAT) correlations were 92 males and 72 females for Study 1 and 102 males and 83 females for Study 2. Sample sizes for the correlations between the Defining Issues Test (DIT) and the Raven test were 140 males and 117 females for Study 1 and 127 males and 113 females for Study 2. DIT-P% = most commonly employed index of moral reasoning assessed by the DIT; SAT-V = SAT Verbal subtest; SAT-M = SAT Math subtest.

sures correlated more highly with the DIT than the noncognitive measures, in all 18 forward-selection stepwise analyses we entered the cognitive measures first. What remained to be learned, however, was whether the noncognitive measures offer any incremental validity to the prediction of the DIT *after* the cognitive measures were partialled. Results revealed that the leisure activity involving video games was the only noncognitive correlate with incremental validity beyond the cognitive variables. Video game activity displayed 5% (Study 1) and 6% (Study 2) incremental validity following SAT-M scores, and 5% (Study 1) and 3% (Study 2) incremental validity following Raven's matrices' scores. No incremental validity, however, was evident when SAT-V scores were included in the analysis.

Discussion

In this study, we found that highly gifted adolescents (top 0.5%) displayed exceptionally high DIT scores relative to their average ability peers and college students four to five years older. This may lead some to suspect that the intellectually gifted are better able to deal with moral issues in a way that is distinct from the general superiority of their intellectual abilities. Yet our evaluation of the hypothesis that the DIT has unique predictive properties relative to markers of general intelligence was evaluated with negative results. This is especially troublesome for establishing the uniqueness of the DIT inasmuch as, given the elite intellectual level of our participants, their restriction of range on the ability measures actually enhances the likelihood of the noncognitive measures achieving incremental validity.

In fact, none of the many noncognitive variables used here manifested significant DIT correlations across both studies for both genders. Moreover, for the gender-mixed samples, the only variable that achieved incremental validity in the prediction of DIT scores was video games (a negative relationship), but only when in direct competition with SAT-M and the Advanced Ra-

ven's matrices; it did not achieve incremental validity following the removal of variance shared with SAT-V. Thus, Lykken (1991) may have hit on something when he chose verbal ability as the critical variable to be partialled from moral reasoning scores. When verbal ability is partialled from the DIT, the remaining variance of the DIT does not overlap significantly with any of the 62 criterion variables examined here. Given that the DIT failed to manifest shared overlap with a wide array of psychological criteria beyond verbal ability, we offer the following generalization: The DIT is simply another way of measuring verbal ability, probably the most salient marker of general intelligence.⁵ Future research with the DIT undoubtedly should be aimed at falsifying this generalization.

Now, to be sure, our results are not definitive as we have in no way exhausted the full range of criterion variables that could be justified in a study of this kind. Other criteria may very well paint a different picture. We believe, however, that we have provided sufficient evidence for the need for such a picture if we are to continue using the DIT in psychological research as a measure of moral reasoning ability and casting it as an instrument in possession of predictive properties distinct from verbal ability.⁶

Finally, although the present investigation focused on the DIT and provided disconfirming evidence for its predictive power distinct from verbal ability, Lykken's (1991) initial query, which motivated this study, is actually more general. Other psychological constructs, especially those of the fulfillment variety (purporting to index sophisticated forms of human development), such as ego development and self-actualization, appear intuitively to be likely candidates for analyses similar to those reported here. In fact, Loevinger (1976) reported correlations between her measure of ego development and general intelligence in the .10–.50 range. It is intriguing to speculate on the extent to which findings based on these instruments (and interpreted in terms of constructs they purport to assess) are actually more centrally related to the construct of general intelligence or the specificity of one of its major markers (e.g., verbal ability). Future investigators are well advised to incorporate markers of general intelligence into their designs to ascertain whether their measures (of constructs purporting to possess some distinctiveness from general intelligence) are getting at anything unique and psychologically meaningful.⁷ This could actually result in a more parsimonious, less redundant collection of scientifically significant constructs and measures in psychological science.

⁵ Actually, it would be useful to conduct operational and constructive replications (Lykken, 1968) of this generalization using other measures of moral reasoning (cf. Bear & Rys, 1994; Speicher, 1994).

⁶ For investigators wishing to embark on such studies, it is imperative that a well-established marker of verbal ability be used. A 10-item vocabulary test is not sufficient. Excellent choices are the SAT-V and ACT Reading Comprehension tests. These instruments have the added advantage of being selection criteria at most universities.

⁷ This is straightforwardly determined by partialing general intelligence from one's measure of interest and then evaluating whether its residual variance shares significant overlap with relevant psychological criteria for which the measure was designed to predict (as done here). Ideally, given the problems in psychological research with Type I error (Meehl, 1990a, 1990b), literal replications of statistically significant

References

- Abel, T. M. (1941). Moral judgments among subnormals. *Journal of Abnormal and Social Psychology*, 36, 378–392.
- Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin*, 102, 3–27.
- Ackerman, P. L. (1988). Determinants of individual differences in skill acquisition: Cognitive processes and information processing. *Journal of Experimental Psychology: General*, 117, 299–329.
- Allport, G. W., Vernon, P. E., & Lindzey, G. (1970). *Manual for the Study of Values: A scale for measuring the dominant interests in personality* (3rd ed.). Boston: Houghton Mifflin.
- Bear, G. G., & Rys, G. S. (1994). Moral reasoning, classroom behavior, and sociometric status among elementary school children. *Developmental Psychology*, 30, 633–638.
- Benbow, C. P. (1988). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. *Behavioral and Brain Sciences*, 11, 169–232.
- Benbow, C. P. (1990). Meeting the needs of gifted students through use of acceleration. In M. C. Wang, M. C. Reynolds, & H. J. Walberg (Eds.), *Handbook of special education* (pp. 23–36). New York: Pergamon Press.
- Benbow, C. P. (1992). Academic achievement in mathematics and science between ages 13 and 23: Are there differences among students in the top one percent of mathematical ability? *Journal of Educational Psychology*, 84, 51–61.
- Benbow, C. P., & Stanley, J. C. (1983). *Academic precocity: Aspects of its development*. Baltimore: Johns Hopkins University Press.
- Betz, N. E., & Fitzgerald, L. F. (1993). Individuality and diversity: Theory and research in counseling psychology. *Annual Review of Psychology*, 44, 343–381.
- Biggs, D. A., & Barnett, R. (1981). Moral judgment development of college students. *Research in Higher Education*, 14, 91–102.
- Boehm, L. D. (1967). Conscience development in mentally retarded adolescents. *Journal of Special Education*, 2, 93–103.
- Brand, C. (1987). The importance of general intelligence. In S. Magil & C. Magil (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 251–265). New York: Falmer Press.
- Carroll, J. B. (1993). *Human cognitive abilities*. Cambridge, England: Cambridge University Press.
- Cauble, M. A. (1976). Formal operations, ego identity, and principled morality: Are they related? *Developmental Psychology*, 12, 363–364.
- Colby, A., & Kohlberg, L. (Eds.) (1987). *The measurement of moral judgment*. New York: Cambridge University Press.
- Dawis, R. V. (1992). The individual differences tradition in counseling research. *Journal of Counseling Psychology*, 39, 7–19.
- Duffy, J. P. (1982). Service programs: Do they make a difference? *Momentum*, 13, 33–35.
- Durkin, D. (1959). Children's concepts of justice: A comparison with the Piaget data. *Child Development*, 30, 59–67.
- Gordon, R. A. (1987). SES versus IQ in the race-IQ delinquency model. *International Journal of Sociology and Social Policy*, 7, 30–96.
- Gough, H. G., & Heilbrun, A. B. (1983). *The Adjective Checklist Manual*. Palo Alto, CA: Consulting Psychologists Press.
- Gustafsson, J. E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 5, pp. 35–71). Hillsdale, NJ: Erlbaum.
- Hanson, R. A., & Mullis, R. L. (1985). Age and gender differences in empathy and moral reasoning among adolescents. *Child Study Journal*, 15, 181–188.
- Humphreys, L. G. (1979). The construct of general intelligence. *Intelligence*, 3, 105–120.
- Humphreys, L. G. (1991). Limited vision in the social sciences. *American Journal of Psychology*, 104, 333–353.
- Humphreys, L. G. (1992). Commentary: What both critics and users of ability tests need to know. *Psychological Science*, 3, 271–274.
- Humphreys, L. G., Lubinski, D., & Yao, G. (1993). Utility of predicting group membership and the role of spatial visualization in becoming an engineer, physical scientist, or artist. *Journal of Applied Psychology*, 78, 250–261.
- Humphreys, L. G., Rich, S. A., & Davey, T. C. (1985). A Piagetian test of general intelligence. *Developmental Psychology*, 21, 872–877.
- Jacobs, J. E., & Eccles, J. S. (1992). The impact of mothers' gender-role stereotypic beliefs on mothers' and children's ability perceptions. *Journal of Personality and Social Psychology*, 63, 932–944.
- Jensen, A. (1973). *Educability and group differences*. New York: Harper & Row.
- Keating, D. P. (1974). The Study of Mathematically Precocious Youth. In J. C. Stanley, D. P. Keating, & L. H. Fox (Eds.), *Mathematical talent: Discoveries, description, and development*. Baltimore: Johns Hopkins University Press.
- Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. Goslin (Ed.), *Handbook of socialization theory and research*. Chicago: Rand McNally.
- Kohlberg, L. (1984). *The psychology of moral development*. San Francisco: Harper & Row.
- Laubscher, S. (1988). *The significance of participating in extramural activities*. Unpublished doctoral dissertation, Indiana University.
- Lockley, O. E. (1976). *Level of moral reasoning and student's choice of terminal and instrumental values*. Unpublished doctoral dissertation, University of Chicago.
- Loevinger, J. (1976). *Ego development: Conceptions and theories*. San Francisco: Jossey-Bass.
- Lubinski, D., & Benbow, C. P. (1992). Gender differences in abilities and preferences among the gifted: Implications for the math/science pipeline. *Current Directions in Psychological Science*, 1, 61–66.
- Lubinski, D., & Benbow, C. P. (1994). The Study of Mathematically Precocious Youth: The first three decades of a planned 50-year study of intellectual talent. In R. F. Subotnik & K. D. Arnold (Eds.), *Beyond Terman: Contemporary longitudinal studies of giftedness and talent* (pp. 255–281). Norwood, NJ: Ablex.
- Lubinski, D., Benbow, C. P., & Sanders, C. E. (1993). Reconceptualizing gender differences in achievement among the gifted. In K. A. Heller, F. J. Monks, & A. H. Passow (Eds.), *International handbook for research on giftedness and talent* (pp. 693–708). Oxford, England: Pergamon Press.
- Lubinski, D., & Dawis, R. V. (1992). Aptitudes, skills, and proficiencies. In M. D. Dunnette & L. M. Hough (Eds.), *The handbook of industrial/organizational psychology* (2nd ed., Vol. 3, pp. 1–59). Palo Alto, CA: Consulting Psychologists Press.
- Lykken, D. (1968). Statistical significance in psychological research. *Psychological Bulletin*, 70, 151–159.
- Lykken, D. T. (1991). What's wrong with psychology anyway? In D. partialled correlations are highly desirable (Lykken, 1968). Actually, the same concern regarding the unique contribution of new psychological measures, over and above preexisting measures, has been expressed in other contexts: for example, in human abilities with respect to the construct of general intelligence (Lubinski & Dawis, 1992; Messick, 1992), in personality assessment with respect to the Big Five (Tellegen, 1993), and in counseling research involving the full spectrum of individual-differences variables (Dawis, 1992). These investigators propose a similar methodology to the one used here for gleaming the unique contribution of innovative measures within these psychological spheres.

- Chicetti & W. Grove (Eds.), *Thinking clearly about psychology* (pp. 3–39). Minneapolis: University of Minnesota Press.
- Meehl, P. E. (1970). Nuisance variables and the ex post facto design. In M. Radner & S. Winokur (Eds.), *Minnesota studies in the philosophy of science* (Vol. 4, pp. 373–402). Minneapolis: University of Minnesota Press.
- Meehl, P. E. (1971a). High school yearbooks: A reply to Schwarz. *Journal of Abnormal Psychology*, 77, 143–148.
- Meehl, P. E. (1971b). Law and the fireside inductions: Some reflections of a clinical psychologist. *Journal of Social Issues*, 27, 65–100.
- Meehl, P. E. (1990a). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant using it. *Psychological Inquiry*, 1, 108–141.
- Meehl, P. E. (1990b). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, 66, 195–244.
- Messick, S. (1992). Multiple intelligences or multilevel intelligences? Selective emphasis on distinctive properties of hierarchy: On Gardner's *Frames of Mind* and Sternberg's *Beyond IQ* in the context of theory and research on the structure of human abilities. *Psychological Inquiry*, 3, 365–384.
- Moos, R. H., & Moos, B. S. (1986). *The Manual for the Family Environment Scale* (2nd ed.). Palo Alto, CA: Consulting Psychologists Press.
- Parikh, B. (1980). Moral judgment development and its relationship to family environmental factors in Indian and American families. *Child Development*, 51, 1030–1039.
- Perry, J. E., & Krebs, D. (1980). Role-taking, moral development, and mental retardation. *Journal of Genetic Psychology*, 136, 95–108.
- Piaget, J. (1932). *The moral judgment of the child*. New York: Free Press.
- Plomin, R., & Bergman, C. S. (1991). The nature of nurture: Genetic influences on "environmental" measures. *Behavioral and Brain Sciences*, 14, 373–427.
- Plomin, R., Reiss, D., Hetherington, E. M., & Howe, G. W. (1994). Nature and nurture: Genetic contributions to measures of the family environment. *Developmental Psychology*, 30, 32–43.
- Raven, J. C., Court, J. H., & Raven, J. (1977). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. London: H. K. Lewis.
- Rest, J. R. (1973). The hierarchical nature of moral judgment. *Journal of Personality*, 41, 86–109.
- Rest, J. R. (1975). Longitudinal study of the Defining Issues Test of moral judgment: A strategy for analyzing developmental change. *Developmental Psychology*, 11, 738–748.
- Rest, J. R. (1976). *Moral judgment related to sample characteristics*. Final report to the National Institute of Mental Health (Report No. 8703). Washington, DC: U.S. Government Printing Office.
- Rest, J. R. (1979a). *Development in judging moral issues*. Minneapolis: University of Minnesota Press.
- Rest, J. R. (1979b). *Revised manual for the Defining Issues Test*. Minneapolis: Moral Research Projects.
- Rest, J. R., Cooper, D., Coder, R., Masanz, J., & Anderson, D. (1974). Judging the important issues in moral dilemmas—an objective measure of development. *Developmental Psychology*, 10, 491–501.
- Rest, J. R., & Narvaez, D. (1994). Summary: What is possible? In J. R. Rest & D. Narvaez (Eds.), *Moral development in the professions: Psychology and applied ethics* (pp. 213–224). Hillsdale, NJ: Erlbaum.
- Scarr, S. (1992). Developmental theories for the 1990's: Development and individual differences. *Child Development*, 63, 1–19.
- Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype → environment effects. *Child Development*, 54, 424–435.
- Schmidt, F. L. (1994). The future of personnel research in the U.S. Army. In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personal selection and classification* (pp. 333–350). Hillsdale, NJ: Erlbaum.
- Simmons, C. H., & Zumpf, C. (1986). The gifted child: Perceived competence, prosocial moral reasoning, and charitable donations. *Journal of Genetic Psychology*, 147, 97–105.
- Snarey, J., Reimer, J., & Kohlberg, L. (1984). The socio-moral development of kibbutz adolescents: A longitudinal, cross-cultural study. *Developmental Psychology*, 21, 3–17.
- Snow, R. E., & Lohman, D. F. (1989). Implication of cognitive psychology for educational measurement. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 263–331). New York: Collier Macmillan.
- Speicher, B. (1994). Family patterns of moral judgment during adolescence and early adulthood. *Developmental Psychology*, 30, 624–632.
- Tellegen, A. (1993). Folk concepts and psychological concepts of personality and personality disorders. *Psychological Inquiry*, 4, 122–130.
- Terman, L. M. (1925). *Genetic studies of genius: Vol. 1. Mental and physical traits of a thousand gifted children*. Stanford, CA: Stanford University Press.
- Vernon, P. E. (1961). *The structure of human abilities* (2nd ed.). London: Methuen.
- Whiteman, P. H., & Kosier, K. P. (1964). Development of children's moralistic judgments. *Child Development*, 35, 843–850.

Received March 16, 1994

Revision received October 31, 1994

Accepted November 21, 1994 ■