Journal of Educational Psychology
2001, Vol. 93, No. 3, 604–614

0022-0663/01/$5.00 DOI: 10.1037//0022-0663.93.3.604

Importance of Assessing Spatial Ability in Intellectually Talented Young Adolescents: A 20-Year Longitudinal Study

Daniel L. Shea, David Lubinski, and Camilla P. Benbow
Vanderbilt University

At age 13, 393 boys and 170 girls scoring at the top 0.5% in general intelligence completed the Scholastic Assessment Test Mathematics (SAT-M) and Verbal (SAT-V) subtests and the Differential Aptitude Test (DAT) Space Relations (SR) and Mechanical Reasoning (MR) subtests. Longitudinal data were collected through follow-up questionnaires completed at ages 18, 23, and 33. Multivariate statistical methods were employed using the SAT-M, SAT-V, and a DAT (SR + MR) composite to predict a series of developmentally sequenced educational–vocational outcomes: (a) favorite and least favorite high school class, (b) undergraduate degree field, (c) graduate degree field, and (d) occupation at age 33. Spatial ability added incremental validity to SAT-M and SAT-V assessments in predicting educational–vocational outcomes over these successive time frames. It appears that spatial ability assessments can complement contemporary talent search procedures. The amount of lost potential for artistic, scientific, and technical disciplines that results from neglecting this critical dimension of nonverbal ideation is discussed.

Theory and practice in gifted education have shifted from an emphasis primarily on general cognitive ability (general intelligence or "g") to an appreciation of the unique information afforded by verbal and quantitative abilities (Benbow & Lubinski, 1996; Benbow, Lubinski, Shea, & Eftekhari-Sanjani, 2000; Benbow & Stanley, 1996; Colangelo & Davis, 1997; Heller, Mönks, & Passow, 1993; Winner, 1996). Attention to these more molecular abilities has proven quite useful for purposes of identification and appropriate developmental placement (i.e., providing educational opportunities commensurate with students' level and pattern of abilities). Both can be measured well early-on through above-level testing, that is, assessing young adolescents with instruments developed on older students (Lubinski & Benbow, 2000). For example, when the Verbal and Mathematics sections of the Scholastic Assessment Test (SAT-V and SAT-M) are administered to 12- and 13-year-olds who have scored in the top 3% on conventional tests routinely administered in their schools, they display essentially the same distributions that high school seniors generate (Assouline & Lupkowski-Shoplik, 1997; Benbow, 1988, 1990; Lubinski & Benbow, 1994). Many of those scoring at or beyond the mean for college-bound seniors can assimilate a full introductory high school course (e.g., chemistry, mathematics, writing) in 3 weeks of full-time study.

The differential and predictive validity of math and verbal reasoning is evinced by comparing students who are relatively more verbally than quantitatively talented. Such students gravitate toward the humanities and social sciences, whereas those with the opposite ability pattern lean more toward engineering and the physical sciences (Achter, Lubinski, Benbow, & Eftekhari-Sanjani, 1999; Lubinski, Webb, Morelock, & Benbow, 2001). Yet, even though the measurement of verbal and quantitative abilities is clearly useful (Lubinski, Benbow, Shea, Eftekhari-Sanjani, & Halvorson, 2001), we suspect that other ability dimensions might provide value-added benefits and, thereby, improve educational counseling and appropriate developmental placement for intellectually talented youth.

Given what is known about the structure and organization of human abilities (Carroll, 1993; Snow & Lohman, 1989), there appears to be at least one dimension of the cognitive spectrum missing, namely, *spatial ability.* Spatial ability (or, more precisely, spatial *visualization*), which dovetails with verbal and quantitative abilities (Gustafsson & Undheim, 1996; Lubinski & Dawis, 1992), represents an "ability in manipulating visual patterns, as indicated by level of difficulty and complexity in visual stimulus material that can be handled successfully, without regard to the speed of task solution" (Carroll, 1993, p. 362). Proficiency in spatial ability has long been associated with success in cognitively demanding educational tracks and occupations such as engineering, architecture, physics, chemistry, and medical surgery, as well as trades such as artisan, certain industrial positions (e.g., die checker, detailer, and pattern checker), surveyor, draftsman, and cartographer (Bingham, 1937; Smith, 1964; Snow & Yalow, 1982; Vernon, 1961). According to Lubinski and Benbow (1992), "although mechanical reasoning and spatial abilities typically are not assessed when selecting individuals for advanced training in basic

Daniel L. Shea, David Lubinski, and Camilla P. Benbow, Department of Psychology and Human Development, Vanderbilt University.

This article is based on a thesis submitted by Daniel L. Shea to Vanderbilt University in partial fulfillment of the masters of science degree. Support for this article was provided by a National Science Foundation Grant MDR 8855625, an anonymous donor, and a Templeton Award for Positive Psychology. An earlier version of this article benefited from comments by Julian C. Stanley and Rose Mary Webb.

Correspondence concerning this article should be addressed to David Lubinski, Department of Psychology and Human Development, Box 512 Peabody Station, Vanderbilt University, Nashville, Tennessee 37203. Electronic mail may be sent to david.lubinski@vanderbilt.edu.

science, strong abilities in these areas are salient characteristics of physical scientists" (p. 63). Gardner (1993) agrees and suggests "it is skill in spatial ability which determines how far one will progress in the sciences" (p. 192). Others have made similar observations (Hedges & Nowell, 1995; Humphreys, Lubinski, & Yao, 1993) and some have launched theoretical explanations (Geary, 1996, 1998; Halpern, 2000; Kimura, 1999). Wide agreement exists that spatial ability distinguishes group membership and performance in certain artistic, engineering, and scientific disciplines.

Why have assessments of spatial abilities been neglected in working with intellectually talented students? This may stem from false beliefs that spatial ability is more relevant to the "vocational trades" than to academic or professional endeavors, inasmuch as the latter tend to place a heavy emphasis on verbal competence (Humphreys & Lubinski, 1996; Lohman, 1994). An alternative possibility, however, is that evidence of the differential and incremental validity of multiple abilities over and above verbal and mathematical abilities (largely reflecting "g") has been lacking (Bennett, Seashore, & Wesman, 1974; McNemar, 1964).

Spatial ability tests display limited usefulness for predicting traditional academic (continuous performance) criteria, partly because most course grades and academic accomplishment assessments are saturated with content specifically indicative of reasoning with numbers and words. If students were required to operate more in complex physical science laboratories, architectural design studios, or in some of the creative arts, there is reason to suspect that measures of spatial ability would contribute to predicting performance and add incremental validity to conjoint verbal and quantitative reasoning assessments (Humphreys et al., 1993; Humphreys & Lubinski, 1996; Lubinski & Dawis, 1992). We also suspect that spatial ability contributes incremental validity to mathematical and verbal reasoning abilities in the prediction of educational settings that students choose to enter as well as migrate from and, hence, aspects of their subsequent career development (Humphreys & Lubinski, 1996; Smith, 1964; Vernon, 1961).

Furthermore, identifying students with exceptional spatial abilities has an important societal function. Some scholars have expressed concern about our failure to identify and foster the development of individuals with nonverbal intellectual gifts and how this does a disservice not only to students, but to society as well (Gohm, Humphreys, & Yao, 1998; Humphreys & Lubinski, 1996; Silverman, 1998; Stanley, 1994). A report by the National Science and Technology Council (2000) on our future scientific, technical, and engineering workforce reads, "If current trends persist, our nation may not have all of the talent it will need to enable the innovation process that has given America a strong economy and high quality of life" (p. 4).

The purpose of this study was to evaluate the potential utility of measuring spatial ability for predicting educational and vocational outcomes through age 33 in an intellectually talented sample identified with mathematical or verbal reasoning measures during the seventh grade. The framework for this longitudinal research extends across four developmentally sequenced outcomes: (a) favorite and least favorite high school class, (b) undergraduate degree major, (c) graduate degree major, and (d) occupation at age 33 years. Descriptive summaries are presented for each outcome and results from a series of multivariate analyses are reported. Two hypotheses are tested. First, that individual differences assessed in adolescence across spatial, verbal, and quantitative reasoning abilities distinguish educational and vocational groups long after their initial assessment. Second, and more specifically, assessing spatial ability provides incremental validity to predictions of educational and vocational outcomes throughout the life span, over and above verbal and quantitative abilities. To the extent that spatial ability is an important predictor of group memberships in a developmentally consistent way, we gain support for augmenting talent identification procedures and refining educational interventions (Benbow & Stanley, 1996) and counseling (Dawis, 1992, 1996) aimed at intellectually talented youth by incorporating spatial ability measures.

Method

Participants

The participants were 170 girls and 393 boys identified between ages 12 to 14 by the Study of Mathematically Precocious Youth as representing approximately the top 0.5% of general ability for their age group (Lubinski & Benbow, 1994). All were identified through talent searches performed in 1976 and 1978 in the mid-Atlantic region and scored at least 500 on the SAT-M or 430 on the SAT-V. (Although these individuals were selected on the basis of either of these cutting scores, the group was first preselected to represent the top 3% in math ability.) Their overall SAT performance was comparable to that of college-bound high school seniors, but was at least 4 years premature. Further assessment was conducted in the spring of the seventh grade, with participants completing a number of supplemental tests and questionnaires. Follow-up questionnaires were mailed to participants at ages 18 (5-year follow-up), 23 (10-year follow-up), and 33 (20-year follow-up). After several reminders, response rates for the respective questionnaires (without deducting from the denominator participants not contacted or whom had died) became: males, 84%, 69%, 78%; females, 84%, 67%, 82%. A small number of participants, 17 males (4.3%) and 10 females (5.9%), were lost track of early in the study and did not participate in any of the follow-ups. Completing all three follow-ups were 220 males (56.0%) and 101 females (59.4%).

Ability Variables

In addition to the SAT subtests used for selection, all participants completed two subtests of the Differential Aptitudes Tests: Mechanical Reasoning (DAT-MR) and Space Relations (DAT-SR). The DAT was designed for use with 8th through 12th graders for purposes of educational and vocational counseling (Bennett et al., 1974). These subtests assess "abilities to recognize everyday physical forces and principles [MR], and to visualize concrete objects and manipulate those visualizations [SR]" (p. 6). The DAT-SR loads primarily on spatial visualization, and the DAT-MR also loads substantially on this factor. Carroll (1993), speaking of the family of mechanical tests, states that "regardless of an individual's experience with mechanical objects, they tap a basic ability in spatial visualization" (p. 324). Combining DAT-SR and DAT-MR also attenuates the problem of *construct irrelevant error* (Cook & Campbell, 1976), because of the relatively high degree of specificity inherent in many spatial tests (Lohman, 1988). Both subtests are timed (DAT-MR, 30 min; DAT-SR, 25 min); most participants finished well within the allotted time. The maximum possible scores are 70 for the DAT-MR and 60 for the DAT-SR.

An equally weighted linear composite of the DAT subtests, *DAT-C*, was created by standardizing each subtest separately ($M = 0$, $SD = 1$), summing the standardized scores, and then recentering this sum around zero. DAT-C represents the measure of spatial ability used here. SAT-M and SAT-V also were standardized around a mean of 0 and standard deviation of 1. All analyses were based on test score data for the entire sample of 563 participants. Reference to these variables throughout is in terms of their standardized values unless otherwise specified.

Comparison Group Variables

Data from the follow-up surveys were used to classify participants into distinct and mutually exclusive groups for ability comparisons. These data include (a) favorite and least favorite high school class, (b) undergraduate degree major, (c) graduate degree major, and (d) occupation at age 33. (Descriptive statistics may be obtained from the authors.) The high school class data came from the 5-year follow-up. Participants were presented a list of 13 options for which they indicated their favorite and least favorite class. Two groups were formed for each item: a humanities–social science group and a math–science group. The humanities–social science group comprised art, English, foreign languages, humanities, music, and social sciences courses. The math–science group comprised computer science, mathematics, natural science, and physical sciences courses. Responses not used included physical education, vocational, and "school in general." The percentage of unused responses for favorite class and least favorite class, respectively, were 13% and 36% for males and 15% and 45% for females.

Data on undergraduate degree majors came primarily from the 20-year follow-up. For nonrespondents to this questionnaire, the 10-year follow-up was consulted. Only data regarding the first conferred degree were used. The nine groups formed (with group *n*s in parentheses) were biology fields (36), business (17), electrical engineering (79), other engineering (73), humanities (66), math–computer science (73), physical sciences (43), and social sciences (54). A small group of participants ($n = 11$) majored in other widely scattered fields and were not coded for this variable. Double majors and two-degree recipients were included only if both majors belonged to the same group, leading to 25 more exclusions. Only 22 participants are known not to have earned a bachelor's degree as of age 33 (males, 4.7%; females, 4.0%). Two engineering groups were formed because the electrical engineering group was large enough to permit this refinement.

Graduate degree information was obtained from the 20-year follow-up or the 10-year follow-up (in that order). Graduate degree majors were grouped (with group *n*s in parentheses) as follows: business (45), electrical engineering (31), other engineering (35), humanities–social sciences (33), law (34), math–computer science (32), medicine (41), and natural–physical sciences (29). Eleven participants earned degrees in areas other than these and were not coded. Only the highest and first graduate degree earned was considered. (For instance, someone earning an M.S., Ph.D., and then an M.D., would be placed in the field related to the PhD.) On the basis of 10-year and 20-year data, 291 participants received graduate degrees (males, 59%; females, 57%) and another 31 were currently attending graduate school (21 males, 5.3%; 10 females, 5.9%).

Occupational information was obtained from the 20-year follow-up. Only participants reporting full-time employment were coded. Occupational groups (with group *n*s in parentheses) were business (68), engineering (87), humanities–social sciences–education (38), law (25), math–computer science (87), medicine (39), and natural–physical sciences (23). This variable proved to be the most difficult to code, largely because of the high degree of specialization among many participants. Use was made of a number of sources in classifying participants into these general categories (Stevens & Hoisington, 1987; U.S. Department of Labor, 1991, 1999). Twenty-eight participants reported occupations that did not fit into any of these categories (e.g., aircraft pilot, administrative assistant, waiter) and were not coded on this variable.

Research Design and Analyses

Educational and occupational groups are described and compared in terms of their distinguishing quantitative (SAT-M), verbal (SAT-V), and spatial (DAT-C) abilities. This represents a *criterion of group membership* methodology, which provides an alternative to traditional criterion-related evidence involving regression of a predictor on some performance measure, for investigating the predictive validity of psychometric tests, but both methods are highly complementary when used together (Humphreys et al., 1993; Rulon, Tiedeman, Tatsuoka, & Langmuir, 1967). For the group membership approach, a variety of statistical methods have been used, including summary statistics (Humphreys et al., 1993), analysis of variance (Govier & Feldman, 1999), hierarchical regression (Wilk, Desmarais, & Sackett, 1995), and discriminant function analysis (Achter et al., 1999; Austin & Hanisch, 1990). Here, groups are compared in terms of trivariate (math × space × verbal) means, or *centroids*, using graphical presentation and a multivariate pairwise test of generalized distances among group means. The test for differences between two centroids is analogous to a *t* test for the significance of a difference between two univariate means. Here, the difference between two trivariate means is the matrix of differences in predictor means separating the groups, adjusted by the covariance matrix using all observations. A useful property of this statistic is that it is adjusted for multicollinearity among the predictor variables. The generalized-distance statistic is compared with an F distribution with k, $N - g - k$ degrees of freedom, where k = the number of nongrouping variables, N = the number of total observations, and g = the number of groups. Results from descriptive and predictive discriminant analyses are reported.

Discriminant function analysis is a multivariate statistical method that examines the possibility of distinguishing among a number of mutually exclusive groups on the basis of a set of continuous variables. It is useful to distinguish between two uses for this method, description and prediction (Huberty, 1994). Descriptive discriminant analysis provides statistics that can assist in making and testing inferences concerning the structural relationship among a set of continuous variables in terms of a categorical criterion variable. Following Stevens (1996, pp. 264–265), statistics used here include (a) the correlations among the discriminant functions and the test variables for giving substantive interpretations of the discriminant functions and (b) the standardized canonical coefficients on the discriminant functions, which suggest the relative importance of each test variable in distinguishing the groups. Another statistical used for evaluating the relative merit of predictor variables is the F-to-remove statistic, obtained from stepwise discriminant analysis (Huberty, 1984). For this statistic, larger F values (and their lower p values) suggest greater importance in achieving group separation.

Departures from multivariate normality can affect results from canonical and parametric discriminant analyses (Mardia, 1971; Stevens, 1996). Hence, all 23 undergraduate, graduate, and occupational group distributions were individually examined for departures from normality using univariate statistics (coefficients of skewness and kurtosis and the Shapiro–Wilk statistic; Stevens, 1996, p. 244), bivariate plots, graphical trivariate representations (using SAS Version 8; SAS Institute, Inc., Cary, NC), and multivariate normality plots (Khattree & Naik, 1999). Several of the group distributions displayed nonnormality, most often in the form of skewness and the presence of outliers. A small number of influential observations were systematically removed to minimize the signs of nonnormality and to ensure that within-group covariance matrices were reasonably uniform. Bartlett's modification of the likelihood ratio test for homogeneity of the within-group covariance matrices (Morrison, 1976; Anderson, 1984) was used iteratively for uniformity appraisals. The test is unbiased (Perlman, 1980) but not robust to nonnormality (in this case making the test conservative). These procedures were deemed necessary to achieve more interpretable statistics derived from the canonical analyses and to make the linear discriminant analyses as powerful as possible. Though tedious, these procedures were successful in ensuring that the distribution for each group was reasonably close to multivariate normal and that their covariance matrices were uniform. The groups from which observations were removed and the small number of participants removed are reported along with the multivariate results.

Predictive discriminant analysis uses a mathematical algorithm to assign probabilities of group membership to each person for all groups under analysis and then assigns each individual to the group for which the highest probability of membership is calculated. Because of differences in group sizes, prior probabilities of group membership were compensated for in the algorithms for all predictive analyses. Both parametric (linear discriminant analysis) and nonparametric (nearest neighbor) analyses were used to evaluate the incremental validity of spatial ability, relative to mathematical and verbal abilities. In multivariate psychologial data, nonnormality is more the rule than the exception, and nonparametric methods, in general, tend to be more powerful than parametric methods when appreciable departures from normality occur (Conover, 1999; National Research Council, 1989). The important statistic in predictive discriminant analysis is the hit rate, or the number of correct classifications divided by the total number of classifications made. Hit rates can be examined for each group or in terms of overall performance, relative to base-rate expectations. In what follows, primary attention is paid to overall performance. Hit rates were used in this study to compare the "full model" using the DAT-C, SAT-M, and SAT-V with a "reduced model" featuring the two SAT subtests alone. No observations were removed before performing the nonparametric analyses.

Results

The results are presented in three parts. First, descriptive statistics are provided on the ability tests. These results are followed by graphical presentation of trivariate (SAT-M × SAT-V × DAT-C) means for the five grouping variables (favorite high school class, least favorite high school class, undergraduate major, graduate school major, and occupation), along with inferential tests on paired comparisons among squared generalized distances between group centroids. Finally, for degree and occupational groupings, results of descriptive and predictive discriminant analyses are reported.

Ability Tests Scores

The means (with standard deviations in parentheses) at ages 12–14 for the unstandardized SAT-M and SAT-V, respectively, were 516 (55) and 467 (70) for girls and 562 (67) and 449 (77) for boys. The corresponding statistics for the DAT-SR and DAT-MR, respectively, were 44.3 (7.6) and 51.0 (6.4) for girls and 45.4 (8.8) and 55.9 (7.4) for boys. Although not selected on measures of spatial ability, these participants performed quite well on the DAT subtests. Median DAT-SR scores (girls, 45; boys, 47) correspond to at least the 95th percentile for eighth graders (the first grade for which norms are available) and above the 80th percentile for twelfth graders (Bennett et al., 1974, pp. 47, 51). For the DAT-MR subtest, the median scores (girls, 52; boys, 57) correspond to at least the 95th percentiles for eighth-grade boys and girls and at least the 90th percentile for twelfth graders (Bennett et al., 1974, pp. 47, 51).

Graphical and Multivariate Group Comparisons

Following a life span development approach, data regarding favorite and least favorite high school class are described first, followed by undergraduate degree field, graduate school degree field, and, finally, occupation at age 33. The first section also provides an explanation of the format of Figures 1–5, which follow a unique method of displaying multivariate data (attributed to Ross, cited in Gower, 1967).

Favorite and least favorite high school class. For these analyses, within-sex standardized means were used across all three abilities. The trivariate means for the humanities–social science and math–science groups for favorite class are displayed in Figure 1. The corresponding statistics for least favorite class are displayed in Figure 2. In these figures, open circles represent the bivariate SAT-M × SAT-V means for each group. The arrows projecting from these circles represent the magnitude (length) and sign (direction) of the mean DAT-C scores. In these and subsequent figures, the three variables are scaled on the same metric, allowing the magnitude of the spatial ability means to be ascertained by direct comparison with the x-axis. Left-facing arrows represent group means less than the grand mean for each sex (based on the data from all participants); right-facing arrows represent group means greater than the grand mean. The distribution of group centroids can be more fully appreciated by mentally rotating the left-facing arrows downward ("beneath" the page) and the right-facing arrows upward ("above" the page) around an imaginary z-axis. This z-axis should be rotated such that it is independent of both x and y (i.e., at a 90° angle to them). Viewed this way, the arrowheads represent the group centroids (or trivariate means), and comparisons among all groups on all three

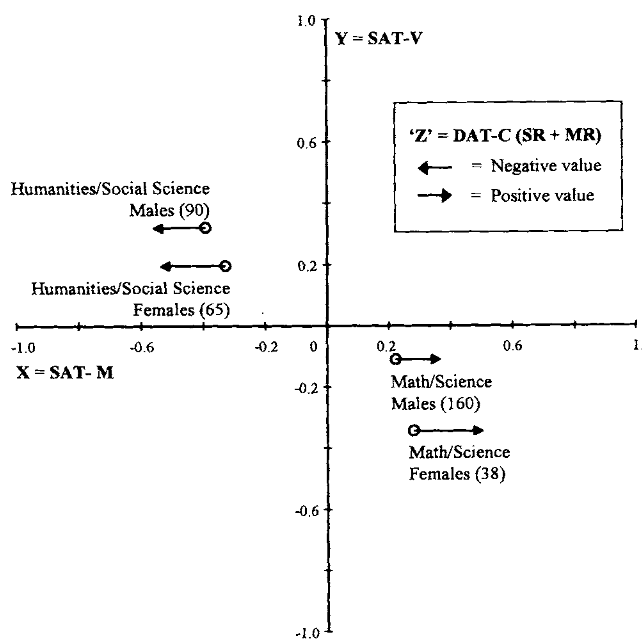

Figure 1. Trivariate means for favorite high school class groups. Ability variables are scaled on a uniform metric. Means are based on within-sex standardization using all participants and are 10% trimmed within groups. Untrimmed group ns are in parentheses. SAT-V = Verbal subtest of Scholastic Assessment Test; SAT-M = Mathematics subtest of Scholastic Assessment Test; DAT-C = Composite of Differential Aptitude Test, representing measure of spatial ability; SR = Space Relations subtest; MR = Mechanical Reasoning subtest.

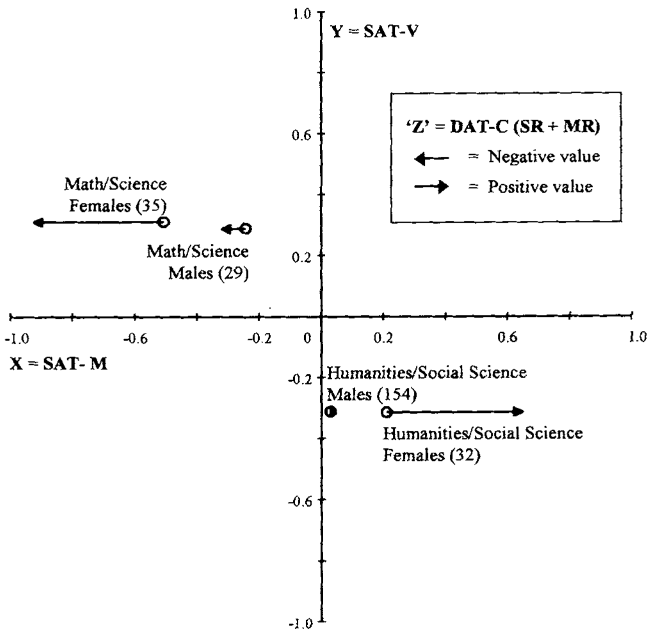

Figure 2. Trivariate means for least favorite high school class groups. Ability variables are scaled on a uniform metric. Means are based on within-sex standardization using all participants and are 10% trimmed within groups. Untrimmed group *ns* are in parentheses. SAT-V = Verbal subtest of Scholastic Assessment Test; SAT-M = Mathematics subtest of Scholastic Assessment Test; DAT-C = Composite of Differential Aptitude Test, representing measure of spatial ability; SR = Space Relations subtest; MR = Mechanical Reasoning subtest.

abilities can be made simultaneously in three-dimensional space. Finally, trimmed means (10%) were employed for displaying these data in order to compensate for a small number of influential data points and a degree of skewness in the group distributions. Trimming does not result in lost information, because the trimmed data still have an attenuated influence on the mean. This method allows the tendency of the majority in a group to be highlighted in the case of slightly nonnormally distributed data (Hoaglin, Mosteller, & Tukey, 1983).

For favorite class (see Figure 1), both boys and girls in the math–science group tended to have strong math and spatial ability and low verbal ability, relative to their gender-equivalent peers who belong to the humanities group. The inverse pattern holds for the humanities–social science groups; both the male and female within-sex standardized means for SAT-M and DAT-C are below those for the entire sample of 563 participants, whereas the SAT-V group means are greater than their gender equivalent grand means. (It is important to bear in mind when comparing these groups that they are all highly able individuals, so differences reflect relative differences in strength.) The finding of generalized distances among group centroid pairs illustrates that both boys and girls who favor math–science classes differ significantly from their same-sex counterparts favoring humanities–social science (boys, $p = .0026$; girls, $p = .0008$). Comparing the male and female groups, the patterns are similar. An important sex difference is that the majority of the boys in this analysis (64%) chose a math or science class as their favorite, whereas a minority of girls (37%) did so.

For least favorite class (Figure 2), the trends evident in the previous figure are generally reversed, but for girls the two centroids are much farther apart than are those for boys. For boys, it appears that mean differences in verbal ability primarily distinguish these math–science and humanities–social science groups, with the latter group having a substantially smaller SAT-V mean but small differences between groups in SAT-M and DAT-C means. The female groups clearly differ on all three means. The math–science least favorite group has lower mathematical and spatial within-sex means and a larger verbal mean, whereas the humanities–social science least favorite group manifests the inverse ability pattern. Here again the difference between the centroids for both male and female groups is significant ($p = .0016$ and $p = .0013$, respectively). A sex difference in percentage of boys and girls selecting between humanities–social science and math–science groups is again evident. Among boys, the majority (84%) chose a humanities or social science class as their least favorite, whereas a minority of girls (48%) did so.

Conferred undergraduate degrees. Shifting focus now from high school to college, Figure 3 displays trivariate means for the eight bachelor's degree groups. Participants who majored in mathematics, computer science, or electrical engineering tend to display tilts favoring mathematical and spatial abilities, whereas those majoring in the humanities, social sciences, and biology appear to be relatively more verbally able. The physical science majors tend to be relatively more able in all three ability domains, while the business majors are relatively less able than all other groups in

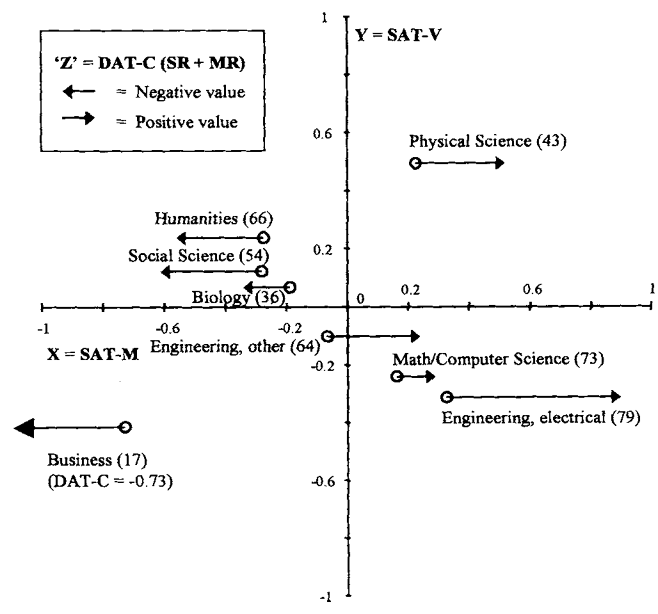

Figure 3. Trivariate means for conferred bachelor's degree groups. Ability variables are scaled on a uniform metric. Means are based on standardization using all participants and are 10% trimmed within groups. Untrimmed group *ns* are in parentheses. SAT-V = Verbal subtest of Scholastic Assessment Test; SAT-M = Mathematics subtest of Scholastic Assessment Test; DAT-C = Composite of Differential Aptitude Test, representing measure of spatial ability; SR = Space Relations subtest; MR = Mechanical Reasoning subtest.

terms of all three abilities. (Note that, for the business group, the large arrowhead illustrates that the magnitude of this group's relative weakness in spatial ability is actually twice as great as that indicated by the displayed length.) The other engineering group seems to be characterized by relatively higher DAT-C scores relative to their SAT-M and SAT-V scores. Overall, the pattern of means is consistent with prior expectations (e.g., engineering requires more spatial ability, whereas business requires relatively little).

With eight groups, there are 28 possible pairwise comparisons between means in terms of their generalized distance. Keeping the experimentwise error rate (alpha level) at .05 means that, using a Bonferroni adjustment, the pairwise test alpha level should be kept at about .05/(8 × 2) = .003. Twleve pairwise comparisons were significant at this level. The electrical engineering group contributed 5 of these. This group was significantly different from biology, business, humanities, physical sciences, and social sciences (all at $p < .0001$). Other significant differences were business versus other engineering, business versus math–computer science, business versus physical science, other engineering versus humanities, other engineering versus social sciences, math–computer science versus humanities, and math–computer science versus social sciences.

Conferred graduate degrees. Trivariate means for the eight graduate degree groups are displayed in Figure 4. As with the undergraduate degree means, these centroids are consistent with prior expectations. The humanities–social science group has the highest SAT-V mean, the math–computer science group has the highest SAT-M mean, and the electrical engineering group has the highest DAT-C mean.

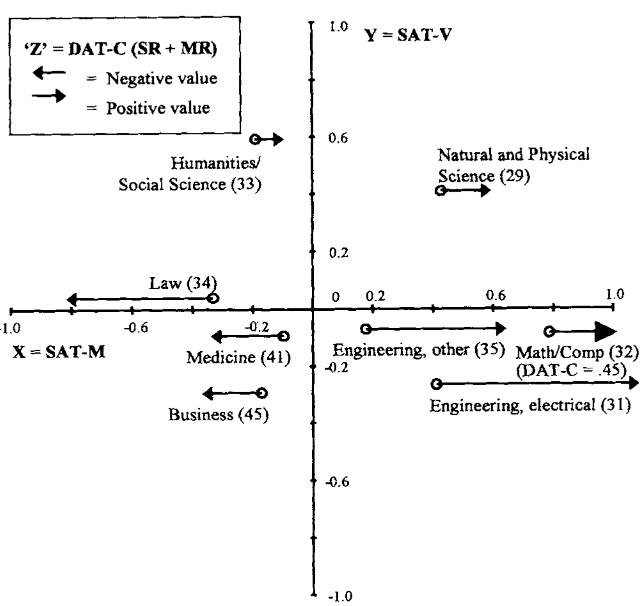

Figure 4. Trivariate means for conveyed graduate degree groups. Ability variables are scaled on a uniform metric. Means are based on standardization using all participants and are 10% trimmed within groups. Untrimmed group *ns* are in parentheses. Comp = computer science; SAT-V = Verbal subtest of Scholastic Assessment Test; SAT-M = Mathematics subtest of Scholastic Assessment Test; DAT-C = Composite of Differential Aptitude Test, representing measure of spatial ability; SR = Space Relations subtest; MR = Mechanical Reasoning subtest.

Upon close inspection, however, it can be seen that consideration of all three abilities, conjointly, greatly facilitates the task of distinguishing group membership. For instance, the math–computer science and other engineering groups can be distinguished only in terms of mathematical ability, whereas spatial ability provides the greatest separation between the other engineering group and medicine. In turn, the medical students' centroid is best distinguished from the humanities–social science group by verbal ability. It also can be seen that the separation between the law group and the electrical engineering group is quite large when all three abilities are considered simultaneously. (Note again the large arrowhead for the math–computer science group—or a spatial ability mean twice the magnitude of that indicated by the displayed length.)

Thirteen paired comparisons of centroids were significant at $p < .003$ (controlling for experimentwise error). The differences (with all p values < .001) were business versus math–computer science, business versus humanities–social sciences, business versus electrical engineering, business versus natural–physical sciences, law versus math–computer science, law versus other engineering, law versus electrical engineering, law versus natural and physical sciences, humanities–social sciences versus math–computer science, humanities–social sciences versus other engineering, humanities–social sciences versus electrical engineering, medicine versus math–computer science, and medicine versus electrical engineering.

Occupational groups. Occupational group data are displayed in Figure 5. Despite significant migration across categories from undergraduate or graduate degree groups to occupational groups (evinced by the changing *ns*), the pattern in previous figures is largely maintained here. Again, the importance of spatial ability is evident from analysis of these data. Comparing the engineering and medicine groups, for instance, the SAT-M provides little differentiation, but the DAT-C and, to a lesser extent, SAT-V seem to account for the separation between these centroids. Likewise, what distinguishes the law group from the business group (in fact from all of these groups) appears to be their relatively lower spatial ability mean. Six of the 21 pairwise comparisons were significant at an alpha level corrected for experimentwise error of .05/(7 × 2) = .0036: engineering versus business, engineering versus law, engineering versus medical, engineering versus humanities–social sciences–education, math–computer science versus business, and math–computer science versus law (all at $p < .001$).

Discriminant Analyses

A series of predictive discriminant analyses (both linear discriminant analyses and nonparametric nearest-neighbor analyses) were executed to ascertain whether spatial ability afforded incremental validity to SAT-M and SAT-V in the prediction of undergraduate and graduate degrees and occupations. Although the primary goal of the following series of analyses was to evaluate whether the incremental validity gleaned by spatial ability was statistically significant, relative to SAT-M and SAT-V, the functions uncovered also were examined for their substantive cohe-

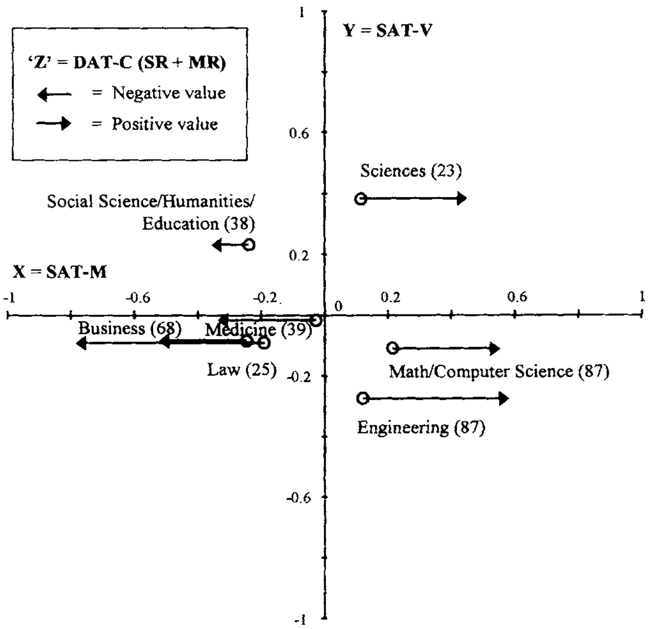

Figure 5. Trivariate means for occupational groups at age 33. Ability variables are scaled on a uniform metric. Means are based on standardization using all participants and are 10% trimmed within groups. Untrimmed group *ns* are in parentheses. SAT-V = Verbal subtest of Scholastic Assessment Test; SAT-M = Mathematics subtest of Scholastic Assessment Test; DAT-C = Composite of Differential Aptitude Test, representing measure of spatial ability; SR = Space Relations subtest; MR = Mechanical Reasoning subtest.

siveness (descriptive discriminant analyses). Yet, with respect to this latter point, it is important to keep in mind that the structural parameters or weights found in discriminant function analyses vary as a function of sample heterogeneity.

Conferred undergraduate degrees. Eight influential observations were removed for the linear model analyses in order to better approximate multivariate normality across the eight degree groups. Observations were removed from the following groups (with *ns* in parentheses): biology (1), electrical engineering (1), other engineering (2), math–computer science (3), and physical sciences (1). The descriptive discriminant (canonical) analysis uncovered two significant discriminant functions: $p = .0001$ for the first function and $p = .02$ for the second. The first function (which accounted for 78% of the between-group variance) runs between (roughly) the humanities group and the electrical engineering group; the second function (which accounted for 18% of the between-group variance) runs from business to the physical sciences group.

These standardized canonical correlations afford substantive interpretation. For the first function, the correlations are SAT-V: −.56, SAT-M: .46, and DAT-C: .76 and represents primarily a bipolar spatial–verbal dimension. For the second function, the correlations are SAT-V: .92, SAT-M: .33, and DAT-C: .21, primarily a verbal dimension.

For the first canonical function, the standardized canonical coefficients (which reveal the relative importance of predictor variables for determining canonical functions after other variables are entered) are SAT-V: −.66, SAT-M: .92, and DAT-C: .96. Hence, SAT-V offers unique information for distinguishing undergraduate groups, over and above the other tests, and the SAT-M and DAT-C both provide unique information, although the redundancy between them (for achieving group discrimination) is indeterminate. For the second function, the standardized canonical coefficients are SAT-V: .75, SAT-M: .32, and DAT-C: .25; the SAT-M and DAT-C are redundant to this function after the SAT-V is considered.

The *F*-to-remove test statistics supports the hypothesis that all three tests add to the predictive efficiency of the discriminant analysis, but the SAT-M appears to add the least information, $F(7, 417) = 2.39$, $p = .021$, whereas the DAT-C, $F(7, 417) = 5.65$, $p = .0001$, and SAT-V, $F(7, 417) = 5.25$, $p = .0001$, contribute fairly equally overall in distinguishing undergraduate degree group membership. Taken together, these structural analyses support the idea that, among the three abilities, the SAT-M provided the least information for predicting individuals' group membership in undergraduate degree areas.

The predictive discriminant analyses also support the hypothesis that spatial ability plays an important role in educational–vocational outcomes. Linear discriminant analysis using all three abilities yielded a hit rate of 26% (or between 8% and 12% above chance expectation). (The 8% value represents the achieved hit rate minus the highest prior probability, whereas the 12% value is calculated as the hit rate minus the sum of squared prior probabilities.) Using only the SAT-V and SAT-M, the hit rate was lower, 22%. However, neither linear discriminant model provided a particularly good fit with the data, although the full model was somewhat better. Over eight groups, the full model resulted in two marginal (group) hit rates of 0% (biology and business), whereas the partial model (SAT-V + SAT-M) resulted in four marginal hit rates of 0% (biology, business, math–computer science, and social science).

The nonparametric (nearest neighbor) predictive discriminant analysis using all three predictors (and with no observations removed) yielded a hit rate of greater than 34%, or 16% to 20% over chance, and resulted in a marginal hit rate of 0% for one of the eight groups (business) and a high hit rate of 47% (humanities). The nearest-neighbor parameter, k, used for this analysis was 4, indicating that classification for each individual was based on information from the four (more in cases of ties) individuals lying closest in terms of generalized (or *Mahalanobis*) distance. In nearest-neighbor methods, the choice of k is usually relatively uncritical (Hand, 1982). For this initial analysis, different values of k were tested against a criterion of "jackknifed" cross-validation error rates. For simplicity, this parameter value was retained for all subsequent nearest-neighbor analyses.

The hit rate using only the SAT-V and SAT-M was slightly lower, 32%, but this reduced model produced somewhat more uniform marginal hit rates, ranging from 12% for the undergraduate business field group to 40% for the math and computer science undergraduate group.

Conferred graduate degrees. Two observations were removed before performing the linear analyses for this criterion variable: one from the other engineering group and one from the humanities–social science group. All three discriminant functions were significant at alpha = .05; the first accounted for 63% of the between-group variance ($p < .0001$), the second accounted for 24% ($p = .0003$), and the third accounted for 12% ($p = .026$). The standardized canonical correlations for the first discriminant function are SAT-V: −.31, SAT-M: .65, and DAT-C: .67. For the second function, the standardized canonical correlations are SAT-V: .97, SAT-M: .09, and DAT-C: .12. For the third function, the standardized canonical correlations are SAT-V: −.05, SAT-M: −.92, and DAT-C: .91. So the first discriminant function is a bipolar verbal–spatial–quantitative factor, the second a purely verbal factor, and the third a bipolar quantitative–spatial factor.

The standardized coefficients for the canonical functions mirror the above pattern of canonical correlations. For the first function, these coefficients are SAT-V: −.16, SAT-M: .77, and DAT-C: .77. For the second, the coefficients are SAT-V: .99, SAT-M: .21, and DAT-C: .27. For the third, the coefficients are SAT-V: −.02, SAT-M: −.60, and DAT-C: .57. The F-to-remove statistics, with p values for $F(7, 268)$ in parentheses, are SAT-V: 3.96 (.0004), SAT-M: 4.04 (.0003), and DAT-C: 4.14 (.0002). Together, these descriptive statistics support the hypothesis that all three tests are important in discriminating among graduate degree groups and that the SAT-M and DAT-C each contribute.

The predictive discriminant analyses support the hypothesis that all three abilities enhance discrimination between graduate degree groups. The chance expected hit rate was between 13% and 16%. For the linear discriminant analyses, the full model (SAT-V + SAT-M + DAT-C) and the reduced model (SAT-V + SAT-M) produced the same overall hit rate, 23%. However, the full model provided better marginal hit rates, ranging from 10% for electrical engineering to 38% for business. The reducing model, on the other hand, produced three degree groups of eight with hit rates of 0% (other engineering, electrical engineering, and law), and a high hit rate of 62% for graduate business majors.

For the nonparametric (nearest-neighbor) discriminant analyses, the full and reduced model again achieved equivalent hit rates, 30%. Again, the graduate-degree-group marginal hit rates were more evenly distributed for the three predictor analysis, ranging from 20% for graduate business majors to 44% for math and computer science majors. The reduced model analysis, using only the SAT subtests, provided a marginal hit-rate range from 3% (law) to 46% (medicine). These linear and nonparametric predictive discriminant analyses imply that the DAT-C provides a useful refinement to the SAT subtests in the task of predicting the graduate degree groups into which intellectually talented students self-select.

Occupational groups. Fourteen participant observations were removed from the linear descriptive analyses. The groups from which observations were removed (with removed *n*s in parentheses) were business (5), math–computer science (4), medicine (3), natural and physical sciences (1), and humanities–social science–education (1). In regard to the canonical correlation analysis, only one function was significant at alpha = .05, accounting for 81% of the between-group variance, $F(18, 970) = 4.14$, $p < .0001$. The canonical correlations reflect a bipolar verbal–spatial factor, SAT-V: −.30, SAT-M: .18, and DAT-C: .93. The test variables' canonical coefficients on this function, however, do not support that the SAT-M and DAT-C possess differential utility for separating occupational groups, SAT-V: −.37, SAT-M: .89, and DAT-C: .99.

For the stepwise discriminant analysis, performed to obtain F-to-remove statistics, the SAT-M failed to enter, hence this test offered no incremental validity relative to SAT-V and DAT-C. Moreover, the DAT-C appears to be the dominant predictor. The F-to-remove statistics are DAT-C, $F(6, 344) = 9.42$, $p < .0001$, and SAT-V, $F(6, 344) = 2.49$, $p = .023$. Hence, the DAT-C accounted for 72% of the explained variance for this reduced (SAT-V + DAT-C) model.

Results from the linear and nonparametric discriminant function analyses, however, did not fully complement the results of these descriptive analyses. For the linear analyses, both the full model (SAT-V + SAT-M + DAT-C) and the reduced model (SAT-V + SAT-M) resulted in hit rates of 29%. Both models also produced similar patterns of marginal hit rates. Two of the seven marginals in the full model were 0% (medicine and natural and physical sciences), and the highest hit rate was 53% (engineering). Three of the seven hit rates were 0% in the reduced model (law, medicine, and natural and physical sciences), with the highest marginal obtained for the math–computer science group (48%).

For the nonparametric (nearest neighbor) predictive discriminant analyses, the reduced model combining the SAT subtests outperformed the full model, 40% versus 37%. Marginal hit rates were largely equivalent for both models: full model range, 12% (law) to 41% (math–computer science) and reduced model range, 12% (law) to 56% (engineering).

These results were surprising, given the superiority of the DAT-C for distinguishing occupational groups in the preceding descriptive analyses. Therefore, additional linear and nonparametric analyses using only the SAT-V and DAT-C were performed for comparison with the reduced model featuring the SAT subtests. For the linear discriminant analysis, the SAT-V + DAT-C model resulted in an overall hit rate of 31% (a 2% increment), with two marginal hit rates of 0% (medicine, and natural and physical science) and a high marginal hit rate of 51% (engineering). The nonparametric SAT-V + DAT-C analysis produced a lower overall hit rate than did the SAT-V + SAT-M model (36% vs. 40%) but produced a similar distribution of marginal hit rates: 15% (medicine) to 49% (engineering).

Discussion

Verbal and quantitative abilities alone do not provide a sufficiently descriptive portrait of the cognitive diversity in intellectually talented students. Individual differences in this special population across verbal, quantitative, *and* spatial abilities around age 13 were conspicuously related to educational and vocational group memberships over 20 years. The life span developmental data displayed in Figures 1–5 vividly highlight distinguishing ability-pattern → group membership relations, and the statistical relations uncovered were robust. Intellectually talented adolescents with stronger spatial ability relative to verbal ability were more likely to be found in engineering and computer science–mathematics fields, whereas those with the inverse ability pattern tended to gravitate toward humanities, social science, organic science, medical arts, and legal fields.

A similar pattern also emerged for relative quantitative versus verbal strengths. However, in this study, spatial ability provided somewhat greater overall discriminative power than quantitative ability. This finding is in line with theories about human abilities that stress the importance of a verbal–spatial bipolar dimension (Eysenck, 1995) or distinct and complementary verbal and spatial dimensions (Eliot, 1987; Lohman, 1994; Vernon, 1961) in addition to general intelligence. Clearly, relative to quantitative and verbal ability, spatial ability provides unique information for predicting the educational–vocational tracks that these students self-selected.

Participants in this sample were preselected to represent the top 3% of mathematical ability. That is, before they took the SAT and then later the spatial ability measures, they had scored in the top 3% on the mathematics section of a standardized achievement test (e.g., Iowa Test of Basic Skills). Therefore, the sample variance on this ability is somewhat restricted. Moreover, the DAT measures were somewhat encumbered by ceiling effects. Of course, these methodological and statistical considerations worked against the hypotheses evaluated here. That is, in spite of these shortcomings, this investigation uncovered a huge range in spatial ability among intellectually gifted students identified by conventional talent-search procedures and documented the educational–vocational implications of these individual differences over a 20-year period.

An issue of particular concern is the likelihood that some intellectually promising students are not being identified by current practices, because of the lack of attention given to spatial ability (Humphreys & Lubinski, 1996; Stanley, 1994). Given the correlational structure for verbal, quantitative, and spatial abilities, there are obviously large numbers of "high-space" (i.e., spatially talented) students who do not meet the minimum math or verbal criteria for participation in talent searches as they are currently performed. Yet they are "at promise" for high achievement in a number of intellectually demanding educational and careers tracks, such as architecture and engineering. Using mathematical, spatial and verbal assessments on a stratified random sample of U.S. high school students (Humphreys et al., 1993), it can be shown that selecting for the top 3% of verbal–mathematical ability will result in the loss of more than half of the students representing the top 1% of spatial ability! This special population needs to be identified for several reasons, including because of increasing importance to our more technologically oriented society. Another added advantage is that spatial visualization manifests a lower correlation with socioeconomic status (SES), relative to mathematical and verbal reasoning abilities (viz., in the low .30s, or .10 correlational units less). Therefore, utilizing spatial ability measures will identify more talented students from lower SES levels than do current talent-search procedures.

How, then, might spatially gifted students be identified? If students already identified by the SAT (or the American College Test [ACT]) and then evaluated for possible academic intervention (i.e., appropriate developmental placement) were routinely assessed on spatial ability, then some spatially talented students who score below the selection criteria on SAT-M or SAT-V (or their ACT counterparts) could be picked up. If the statistics just presented are accurate, however, then a large number still will remain lost—those who do not meet the cutting scores for inclusion in talent searches. An approach likely to achieve success is for educators and counselors to become more aware of nontest signs of spatial giftedness: patterns of grades favoring science classes and labs, math classes, and vocational courses; levels of achievement and interests in hobbies requiring building, repairing, or creating; interests in "things" (rather than "people" or "ideas") and in tinkering with objects, and preference for reading science fiction over nonfiction (Gohm et al., 1998; Humphreys et al., 1993; Prediger, 1976). Such activities are signs of exceptional nonverbal abilities. We suggest that these students also should be considered for spatial ability assessments and, if they distinguish themselves on these measures, be given challenging educational opportunities for developing their abilities.

In addition to identification and counseling practices, the school curriculum might need to be adjusted to accommodate these students and help them use their potential. Schools are so verbal in orientation that a great many high-space kids are likely to feel uncomfortable in them and, perhaps, perceive themselves as not "college material." Students with exceptional strength in spatial ability should be encouraged to explore career options in fields where a premium is placed on nonverbal ideation and involves more hands-on activities: architecture, engineering, physical sciences, technical disciplines, as well as a number of the creative arts and vocational fields.

Several investigators have recognized the educational–vocational implications of spatial ability for high school students and young adults (Halpern, 1997, 2000; Hedges & Nowell, 1995; Lohman, 1988; Neisser et al., 1996; Stanley, 1994). This study supports the idea that individual differences in this attribute are important for identifying, educating, and counseling intellectually precocious young adolescents. If our education system is to fulfill its obligation to both the students in its charge and the society it serves, then we can no longer afford to neglect this crucial feature of the human cognitive repertoire.

References

Achter, J. A., Lubinski, D., Benbow, C. P., & Eftekhari-Sanjani, H. (1999). Assessing vocational preferences among intellectually gifted adolescents adds incremental validity to abilities: A discriminant analysis of educational outcomes over a 10-year interval. *Journal of Educational Psychology, 91,* 777–786.

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis* (2nd ed.). New York: Wiley.

Assouline, S. G., & Lupkowski-Shoplik, A. E. (1997). Talent searches: A model for the discovery and development of academic talent. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (2nd ed., pp. 170–179). Boston: Allyn & Bacon.

Austin, J. T., & Hanisch, K. A. (1990). Occupational attainment as a function of abilities and interests: A longitudinal analysis using Project TALENT data. *Journal of Applied Psychology, 75,* 77–86.

Benbow, C. P. (1988). Sex differences in mathematical reasoning ability among the intellectually talented: Their characterization, consequences, and possible causes. *Behavioral and Brain Sciences, 11,* 169–232.

Benbow, C. P. (1990). Sex differences in mathematical reasoning ability among the intellectually talented: Further thoughts. *Behavioral and Brain Sciences, 13,* 196–198.

Benbow, C. P., & Lubinski, D. (Eds.). (1996). *Intellectual talent.* Baltimore: Johns Hopkins University Press.

Benbow, C. P., Lubinski, D., Shea, D. L., & Eftekhari-Sanjani, H. (2000). Sex differences in mathematical reasoning: Their status 20 years later. *Psychological Science, 11,* 474–480.

Benbow, C. P., & Stanley, J. C. (1996). Inequity in equity: How "equity" can lead to inequity for high-potential students. *Psychology, Public Policy, and Law, 2,* 249–292.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1974). *Manual for the Differential Aptitude Tests* (5th ed.). New York: Psychological Corporation.

Bingham, W. V. D. (1937). *Aptitudes and aptitude testing.* New York: Harper & Brothers.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge, England: Cambridge University Press.

Colangelo, N., & Davis, G. A. (Eds.). (1997). *Handbook of gifted education* (2nd ed.). Boston: Allyn & Bacon.

Conover, W. J. (1999). *Practical nonparametric statistics* (3rd ed.). New York: Wiley.

Cook, T. D., & Campbell, D. T. (1976). The design and conduct of quasi-experiments and true experiments in field settings. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (1st ed., pp. 223–326). Chicago: Rand McNally.

Dawis, R. V. (1992). The individual differences tradition in counseling psychology. *Journal of Counseling Psychology, 39,* 7–19.

Dawis, R. V. (1996). Vocational psychology, vocational adjustment, and the workforce: Some familiar and unanticipated consequences. *Psychology, Public Policy, and Law, 2,* 229–248.

Eliot, J. (1987). *Models of psychological space: Psychometric, developmental, and experimental approaches.* New York: Springer-Verlag.

Eysenck, H. J. (1995). *Genius.* Cambridge, England: Cambridge University Press.

Gardner, H. (1993). *Frames of mind: The theory of multiple intelligences* (10th anniversary ed.). New York: Basic Books.

Geary, D. C. (1996). International differences in mathematical achievement: Their nature, causes, and consequences. *Current Directions in Psychological Science, 5,* 133–137.

Geary, D. C. (1998). *Male, female: The evolution of human sex differences.* Washington, DC: American Psychological Association.

Gohm, C. L., Humphreys, L. G., & Yao, G. (1998). Underachievement among spatially gifted students. *American Educational Research Journal, 35,* 515–531.

Govier, E., & Feldman, J. (1999). Occupational choice and patterns of cognitive abilities. *British Journal of Psychology, 90,* 99–108.

Gower, J. C. (1967). Multivariate analysis and multivariate geometry. *The Statistician, 17,* 13–28.

Gustafsson, J.-E., & Undheim, J. O. (1996). Individual differences in cognitive functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 186–242). New York: Simon & Schuster.

Halpern, D. F. (1997). Sex differences in intelligence: Implications for education. *American Psychologist, 52,* 1091–1102.

Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3rd ed.). Mahwah, NJ: Erlbaum.

Hand, D. J. (1982). *Kernel discriminant analysis.* New York: Research Studies Press.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high scoring individuals. *Science, 269,* 41–46.

Heller, K. A., Mönks, F. J., & Passow, A. H. (Eds.). (1993). *International handbook of research and development of giftedness and talent.* Oxford, England: Pergamon Press.

Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (1983). *Understanding robust and exploratory data analysis.* New York: Wiley.

Huberty, C. J. (1984). Issues in the use and interpretation of discriminant analysis. *Psychological Bulletin, 95,* 156–171.

Huberty, C. J. (1994). *Applied discriminant analysis.* New York: Wiley.

Humphreys, L. G., & Lubinski, D. (1996). Assessing spatial visualization: An underappreciated ability for many school and work settings. In C. P. Benbow & D. Lubinski (Eds.), *Intellectual talent: Psychometric and social issues.* Baltimore: Johns Hopkins University Press.

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.

Khattree, R., & Naik, D. N. (1999). *Applied multivariate statistics with SAS software* (2nd ed.). Cary, NC: SAS Institute.

Kimura, D. (1999). *Sex and cognition.* Cambridge, MA: Massachusetts Institute of Technology Press.

Lohman, D. F. (1988). Spatial abilities as traits, processes, and knowledge. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 181–248). Hillsdale, NJ: Erlbaum.

Lohman, D. F. (1994). Spatially gifted, verbally inconvenienced. In N. Colangelo, S. G. Assouline, & D. L. Ambroson (Eds.), *Jocelyn Wallace National Research Symposium on Talent Development* (2nd ed., pp. 251–264). Dayton: Ohio Psychology Press.

Lubinski, D., & Benbow, C. P. (1992). Gender differences in ability 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. (2000). States of excellence. *American Psychologist, 55,* 137–150.

Lubinski, D., Benbow, C. P., Shea, D. L., Eftekhari-Sanjani, H., & Halvorson, M. B. J. (2001). Men and women at promise for scientific excellence: Similarity not dissimilarity. *Psychological Science, 12,* 309–317.

Lubinski, D., & Dawis, R. V. (1992). Aptitudes, skills and proficiencies. In M. D. Dunnette & L. M. Hough (Eds.), *The handbook of industrial/organizational psychology* (pp. 1–59). Palo Alto, CA: Consulting Psychologists Press.

Lubinski, D., Webb, R. M., Morelock, M. J., & Benbow, C. P. (2001). Top 1 in 10,000: A 10-year follow-up of the profoundly gifted. *Journal of Applied Psychology, 86,* 718–729.

Mardia, K. V. (1971). The effect of nonnormality on some multivariate tests and robustness to nonnormality in the linear model. *Biometrica, 58,* 105–121.

McNemar, Q. (1964). Lost: Our intelligence? Why? *American Psychologist, 19,* 871–882.

Morrison, D. F. (1976). *Multivariate statistical methods.* New York: McGraw-Hill.

National Research Council. (1989). Discriminant analysis and clustering. *Statistical Science, 4,* 34–69.

National Science and Technology Council. (2000, April). *Ensuring a strong U.S. scientific, technical, and engineering workforce in the 21st century* [on-line report.]. Retrieved 2000 from the World Wide Web: http://www.whitehouse.gov/WH/EOP/OSTP/html/workforcerpt.pdf

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77–101.

Perlman, M. D. (1980). Unbiasedness of the likelihood ratio tests for equality of several covariance matrices and equality of several multivariate normal populations. *Annals of Statistics, 8,* 247–263.

Prediger, D. J. (1976). A world-of-work map for career exploration. *Vocational Guidance Quarterly, 24,* 198–208.

Rulon, P. J., Tiedeman, D. V., Tatsuoka, M. M., & Langmuir, C. R. (1967). *Multivariate statistics for personnel classification.* New York: Wiley.

Silverman, L. K. (1998). Personality and learning styles of gifted children. In J. VanTassel-Baska (Ed.), *Excellence in education: Gifted and talented learners.* Denver, CO: Love.

Smith, I. M. (1964). *Spatial ability: Its educational and social significance.* San Diego: Knapp.

Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. Linn (Ed.), *Educational measurement* (pp. 263–331). New York: Collier.

Snow, R. E., & Yalow, E. (1982). Education and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 493–585). New York: Cambridge University Press.

Stanley, J. C. (1994). Mechanical aptitude: Neglected undergirding of technological expertise. *The Journal Portfolio* (Article 7). Evanston, IL: Illinois Association for Gifted Children.

Stevens, G., & Hoisington, E. (1987). Occupational prestige and the 1980 U.S. labor force. *Social Science Research, 16,* 74–105.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

U.S. Department of Labor. (1991). *Dictionary of occupational titles* (4th ed., rev.). Washington, DC: U.S. Government Printing Office.

U.S. Department of Labor. (1999). *O*NET 98: Keeping pace with today's changing workplace.* Washington, DC: U.S. Government Printing Office.

Vernon, P. E. (1961). *The structure of human abilities* (2nd ed.). London: Methuen.

Wilk, S. L., Desmarais, L. B., & Sackett, P. R. (1995). Gravitation to jobs commensurate with ability: Longitudinal and cross-sectional tests. *Journal of Applied Psychology, 80,* 79–85.

Winner, E. (1996). *Gifted children: Myths and realities.* New York: Basic Books.

Received August 16, 2000
Revision received January 2, 2001
Accepted January 2, 2001 ■