Journal of Applied Psychology
1993, Vol. 78, No. 2, 250-261

0021-9010/93/$3.00

Utility of Predicting Group Membership and the Role of Spatial Visualization in Becoming an Engineer, Physical Scientist, or Artist

Lloyd G. Humphreys, David Lubinski, and Grace Yao

This article has two themes: First, we explicate how the prediction of group membership can augment test validation designs restricted to prediction of individual differences in criterion performance. Second, we illustrate the utility of this methodology by documenting the importance of spatial visualization for becoming an engineer, physical scientist, or artist. This involved various longitudinal analyses on a sample of 400,000 high school students tracked after 11 years following their high school graduation. The predictive validities of Spatial–Math and Verbal–Math ability composites were established by successfully differentiating a variety of educational and occupational groups. One implication of our findings is that physical science and engineering disciplines appear to be losing many talented persons by restricting assessment to conventional mathematical and verbal abilities, such as those of the Scholastic Aptitude Test (SAT) and the Graduate Record Examination (GRE).

The *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association, & National Council on Measurement in Education, 1985) discusses criterion-related evidence and lists specific standards solely in terms of relations between individual differences on predictor tests and in criterion performance. Consider, for example, the following quotation that was presented by the authors as the central question in criterion-related validation: "How accurately can criterion performance be predicted from scores on the tests?" (p. 11). It is not obvious to us, however, that this is the only question or always the most important question. We argue that the prediction of group membership also is a desirable criterion for consideration in predictive validation. To illustrate the importance of predicting group membership, we exemplify how spatial ability tests are useful in predicting group membership in various engineering and physical science educational/occupational categories. This empirical demonstration not only reveals the usefulness of this methodology but also has implications for identifying individuals, at both the undergraduate and graduate level, with exceptional talent for and commitment to engineering and physical science disciplines, the second objective of this article. Before proceeding, however, a brief review of the traditional form of predictive validation is necessary to point out its distinctiveness from the group membership approach.

Regressions of criterion performance on predictors may be considered the "classic" approach. Although this approach has much to offer, it also has numerous problems associated with it, as illustrated in the discussion in *Standards* (AERA, 1985) concerning the evaluation of criteria documents. Research in both civilian occupations and military assignments, extending over many years, points clearly to two concerns: First, any one criterion measure contains a substantial quantity of unique variance, and a composite of several measures of performance having widely varying methods variance components is likely to be the most valid (Carroll, 1985; Humphreys, 1985; Lubinski & Dawis, 1992). A second problem with the classic approach to predictive validation arises from the instability of individual differences in performance over successive occasions of measurement during training, from training to performance on the job, and over occasions on the job. Hulin, Henry, and Noon (1990) have recently reviewed this literature, which led to the following question: How many different time periods between testing on predictors and obtaining criterion measures are required in studies of predictive validity? We really do not know.

An Alternative Approach: Predicting Group Membership

The Group Membership Criterion

An alternative to the classic approach is to relate examinees' test scores on predictor tests to the mean obtained by both successful and satisfied members of an existing group. This should not be seen as merely a last resort when a measure of criterion performance for a group is lacking. Valid tests designed for predicting performance criteria may not predict group membership well. This is because, in part, membership in an educational or occupational group is, in a sense, a truly aggregate criterion. Persons in the group have survived institu-

Lloyd G. Humphreys and Grace Yao, Department of Psychology, University of Illinois, Urbana-Champaign; David Lubinski, Department of Psychology, Iowa State University.

Support for this research was provided by the Research Board, University of Illinois, Urbana-Champaign; the Pioneer Fund; and by a grant from the National Science Foundation (MDR8855625).

Appreciation is expressed to Camilla P. Benbow, whose helpful comments improved an earlier version of this article.

Correspondence concerning this article should be addressed to Lloyd G. Humphreys, Psychology Building, University of Illinois, Champaign, Illinois 61820, or to David Lubinski. Requests for reprints should be sent to David Lubinski, Department of Psychology, Iowa State University, Ames, Iowa 50011-3180.

tionally imposed hurdles and, in addition, have become members as a function of a series of their own choices.

The proposal to use group membership as a criterion measure for validation research is not, by any means, new. Yet it is indeed different from the regression of criterion performance on predictors; it has just not typically been used for that purpose. Rulon, Tiedmen, Tatsuoka, and Langmuie (1967), for example, not only presented the methodology in detail but they furnished examples of how it could be used in selection. (Classification of military personnel into assignments constitutes a prime example of the near identity of selection and classification decisions.) Furthermore, their quotation, from an article by Kelley (1940), indicates both the age of the concept and its distinctiveness from standard predictive validation: "If one's profile is well above that of the average participant in a job, it is presumptive evidence that he is fitted for a different and more advanced job" (p. 29). It is not coincidental that Rulon studied under Kelley for his Stanford master's degree.

Establishing Criterion Groups

The selection of individuals when group membership is used as a criterion should be done as carefully as sample selection in classic regression research. The groups should be composed, as described previously, of persons who, on average, are successful and satisfied in their assignment or occupation. The satisfaction requirement can be met by evaluating the length of time members have been in an assignment or occupation, whereas the successful requirement can be met by the time in an assignment or occupation that has acceptable standards of admission and retention. The effect on test scores of experience as a member of the group confounds interpretation but can be avoided by the predictive design. Moreover, the selection decisions that formed the group should not insert invalid dimensions of differences that can be created, for example, by gender, race, or ethnicity. The predictors also should be predictively valid and unbiased, in the regression sense, at the time the group is initially formed.

Methodologies in the Prediction of Group Membership

When investigators have numerous predictors and three or more groups, two appropriate methodologies for predicting group membership are multiple discriminant function and canonical correlation (Tatsuoka, 1988). The former was the focus of Rulon et al. (1967) and was described as a methodology for classifying personnel. It requires measures of n attributes for each member of m existing groups. The aim of the analysis is to determine the accuracy with which examinees have been sorted into groups. Accuracy (validity) is evaluated in terms of homogeneity of individual differences within groups and heterogeneity of means between groups. When the attributes are measured before the formation of the groups, the measures of the attributes may accurately be described as predictors and the validation as prediction of group membership. If the members of the groups are, on average, successful and satisfied, research that allows applied psychologists to predict group membership has obvious implications for practice in vocational counseling (Dawis & Lofquist, 1984; Lofquist & Dawis, 1991), as well as personnel selection and classification.

There are actually some examples in the literature of the use of discriminant function methodology to study the relation of earlier test data to later group membership, but the findings have not been applied to the selection of persons in new samples. For example, Austin and Hanisch (1990) used ability tests and self-report interest questionnaires to predict occupational groups in a gender-mixed sample of 10th-grade students from Project TALENT. (Gender was entered as a predictor.) Accuracy of prediction over 13 years varied widely from group to group and varied as a function of the a priori selection of the specific occupational groups that formed the criterion categories. Five discriminant functions were extracted and interpreted. Verbal and mathematical tests dominated the first function, whereas mechanical, spatial, and mathematical tests dominated the second function, which also accounted for an appreciable proportion of variance. Dimensions of various vocational interests were the primary constituents of three small functions that were also included in their interpretations.

Lunneborg and Lunneborg (1975), moreover, used ability and self-report instruments at the time of college entrance to predict membership in senior major groups. In this restricted range of talent, the first large discriminant function was primarily defined by tests measuring mathematical, mechanical, and spatial abilities. (Verbal abilities and technical-scientific interests have relatively small correlations with this function.) A second discriminant function accounted for substantially less variance and was defined by interest dimensions in a bipolar fashion. Vocabulary and other interest dimensions defined a third, still smaller interpretable discriminant function.

The results of these studies can be interpreted in terms of Vernon's (1950) hierarchical model of intelligence, which is currently receiving a great deal of both applied and theoretical attention (Lubinski & Dawis, 1992); Vernon's ideas are especially relevant to the substantive feature of our study, concerning the selection of students for engineering and physical science disciplines. In this model, general intelligence (g) is viewed as the central dimension common to all cognitive tests, whereas two major group factors, *verbal-numerical-educational* (v:ed) and *practical-mechanical-spatial* (k:m), are more content-saturated with abilities involving linguistic/numerical symbols versus ideation about mechanical/spatial things, respectively. Students were oriented in different directions educationally and occupationally by their level and pattern of abilities in both Austin and Hanisch (1990) and Lunneborg and Lunneborg (1975). Differences in level are related primarily to the general factor, whereas differences in pattern depend on the two major group factors.

Other findings mirroring this pattern of results and of importance for this study include the incidental observation, based on Project TALENT data (Flanagan et al., 1962), in Humphreys, Davey, and Kashima (1986). They used composites designed to measure the general factor and Vernon's two major group factors. They reported that engineers and physical scientists had approximately the same level of high scores on experimentally independent measures of the general factor and on the practical-mechanical–spatial major group factor; their abilities were assessed in high school and their occupational status was determined through longitudinal tracking 11 years following their high school graduation. High school graduates in the 1960s were not screened for college selection on entrance examinations or high school grades having appreciable mechanical-spatial content, nor are they today. The conclusion is inescapable that engineers and physical scientists in Humphreys et al. (1986) had been self-selected for mechanical-spatial abilities.

Additional background for the importance and differential validity of Vernon's major group factors is the long history in military personnel research of finding useful regressions of many criterion measures on spatial visualization tests. Such tests are key components in selection composites used in assigning persons to complex specialities that involve technical problem solving (Humphreys, 1962, 1985, 1986; McHenry, Hough, Toquam, Hanson, & Ashworth, 1990; Smith, 1964; Vernon, 1947). Spatial visualization tests may have had greater opportunity to manifest differential validity for high level technical training in the military than in postsecondary civilian education because the former involves more "hands-on" and fewer "bookish" classroom hours (Lubinski & Dawis, 1992).

In the research that follows, we study in greater depth the educational and occupational significance of the spatial/verbal distinction as we reveal the usefulness of the group membership methodology. The specific research question to be addressed is as follows: Given that selection for undergraduate and graduate education typically restricts ability assessment to mathematical and verbal tests, does this practice result in a loss of talent for scientific and engineering disciplines? We shall examine this question by using group membership as the criterion measure. Our design consists of a simple approximation of the more complex methodologies of canonical correlation and discriminant function analyses.

Method

Source of Data

Subjects were obtained from the Project TALENT Data Bank maintained by the American Institutes for Research, in Palo Alto, California. The test data were gathered in 1960 from a stratified random sample of more than 900 of the nation's high schools (Flanagan et al., 1962). Students in grades 9 through 12 were tested in the sample of schools, for a total sample size of approximately 400,000 (approximately 50,000 students of each gender for each grade). A large number of ability measures (e.g., quantitative, spatial visualization, and verbal) and information tests (concerning specific content domains: art, biology, engineering, journalism, literature, physics, etc.) were administered, along with several self-report scales assessing attitudes, interest, and personality traits. In addition, each student filled out a 398-item autobiographical questionnaire concerning family, school and work experience, hobbies, and health. Test booklets were administered to students over a 1-week interval.[1]

Although the data were collected in the 1960s, no adequate data have been gathered since that lends itself to our research question. It is important to realize, however, that *structural relations* among psychological variables are highly robust to societal changes over time and to cultural differences within society at a given point of time (Lubinski & Humphreys, 1990a). Moreover, means on questionnaires of attitudes and interests are less robust than means on tests of maximum performance (Cronbach, 1990; Willerman, 1979). Thus, even though somewhat more women are entering engineering and the physical sciences today than in the early 1960s, in part as a result of increases in means in relevant interests, attitudes, and perhaps abilities, the kinds of women doing so have probably changed little. That is, the structural relations among these attributes and their relations with occupational choices probably have not changed.

The selection composites. Two broad composites were assembled for each gender and in each of the four high school grades. A Verbal–Mathematics composite was designed to be approximately equivalent to the Scholastic Aptitude Test (SAT). A Spatial-Mathematics composite was designed to be an alternative to the present tests available in the SAT for selecting students in engineering and physical science. Mathematics was included in both composites to ensure realistic selection instruments. Thus, possible differential effects of verbal and spatial abilities can be observed because mathematics was held approximately constant.

The Mathematics, Spatial, and Verbal composites were each formed from several different tests in TALENT to increase the reliability and construct validity of the abilities we wished to measure. These constituent tests, with number of items and raw score weights given in parentheses, follow. They are described more fully in Wise, McLaughlin, and Steel (1979).

The Verbal composite was made up of three tests. The vocabulary test (30 items, raw score weight = 2.5) assessed general knowledge of word items. The English composite (113 items, raw score weight = 1.0) assessed spelling, capitalization, punctuation, usage, and effective expression. The Reading Comprehension test (48 items, raw score weight = 1.25) assessed comprehension of written text across a broad range of topics.

The Spatial composite was made up four tests. The 2-D Spatial Visualization test (24 items, raw score weight = 1.0) assessed the ability to visualize two-dimensional figures when they were turned around or turned over on a flat surface. The 3-D Spatial Visualization test (16 items, raw score weight = 3.0) assessed the ability to visualize two-dimensional figures after they had been folded into three-dimensional figures. The Mechanical Reasoning test (20 items, raw score weight = 1.5) measured deductions based on primitive mechanisms (e.g., gears, pulleys, and springs) and knowledge of the effects of common physical forces (e.g., gravity). Abstract Reasoning (15 items, raw score weight = 2.0) was a nonverbal test of logical relationships in complex figural patterns.

The Mathematics composite consisted of three tests. The Mathematics Information test (23 items, raw score weight = .55) assessed the vocabulary of mathematical notation and definitions. The Arithmetic Reasoning test (16 items, raw score weight = 1.0) assessed the reasoning required to solve common arithmetic problems. Finally, the Introductory Mathematics test (24 items, raw score weight = .55) assessed all forms of mathematics taught through the ninth grade.

Humphreys (1991) has estimated that the reliabilities of these, or very similar, composites are approximately .90. These values were based on conservative estimates of parallel form reliabilities of the components. The construct validity of the composites depends on the test construction methodology found in Humphreys (1985), namely, *systematic heterogeneity.* One selects components that differ as much as possible from each other yet also measure the attribute desired by the investigator. A well-designed composite measure, however, assesses the common attribute more validly than any one component, and scores are less affected by unwanted unique variance. The relevant

[1] These instruments may be obtained through the American Institutes for Research, Palo Alto, California. Most of the questionnaire and test items may be found in two publications: *Project TALENT Mastertape Formats* (American Institutes for Research, 1979) and *The Project TALENT Data Bank* (Wise, McLaughlin, & Steel, 1979).

statistical principle is that the weight of a component in a composite is a direct function of the root of the sum of its variance and its $n - 1$ covariances. If this sum is equalized for all components, they are equally weighted. However, attention was also paid to knowledge of what the components measure psychologically. One source of evidence is the measure's loading on the factor common to the set of proposed components. Another source in our situation was the opportunity provided by the high school curriculum to acquire competence on a proposed component. Thus, for example, arithmetic reasoning was given greater weight than either of the narrower mathematics tests, but not their sum, because the former test typically has greater reasoning variance and is less sensitive to variations in high school curricula.

Interest measures. TALENT's interest scales covered the full range of Holland's (1973) six themes of vocational interests (although some themes were indexed more comprehensively than others). The 17 interest measures, grouped according to Holland's system, were as follows: Physical Science, Biological/Medical Science (Investigative), Literature, Art, Music (Artistic), Social Service (Social), Business Management, Public Service, Sales (Enterprising), Computation, Office Work (Conventional), Mechanical/Technical, Skilled Trades, Farming, Labor, Sports, Hunting/Fishing (Realistic). A complete description of each may be found in Wise et al. (1979, p. 22). All 17 scales were multiplied by appropriate constants to achieve a maximum score of 40.

The occupational interest tests administered in Project TALENT did not include subtle or indirect items. The items measure the examinee's attitudes toward specific concrete targets. The names assigned to the various tests directly reflect the content analyses of the relatively small subsets of highly homogeneous items. Reliabilities were not reported by Project TALENT staff, but it is well known that coefficients of homogeneity (lower-bound estimates of reliabilities) are quite high for relatively small subsets of attitude items that are highly homogeneous in content. There is little information about the stabilities of such items over extended intervals of time. If group discrimination is found in a longitudinal design, however, an unknown but nonzero degree of reliability and stability of each measure exhibiting discrimination is documented.

Biographical data. Subjects in Project TALENT also completed a 398-item autobiographical self-report questionnaire. (This entire questionnaire may be found on pp. 156–167 of *Project TALENT Masterfile Tape Formats;* American Institutes for Research, 1979). From this inventory, several composites were assembled by TALENT (on the basis of student's high school and home experiences). We selected for analysis TALENT's Socioeconomic Index (SES), High School Grades, Academic Courses (taken in high school), High School Guidance, Guidance Elsewhere, Study Habits, Self-Perception of Writing Skill, Self-Perception of Reading Skill, Extra Reading, Variety of Hobbies, Participation in Sports, and Leadership Roles.

Longitudinal data. Project TALENT also contains longitudinal data collected at three time points: 1, 5, and 11 years following high school graduation. For this study, we focused on TALENT's 11-year follow-up data, which described educational and career achievements (e.g., highest degree earned, undergraduate and graduate major, and occupation).

Membership in an occupational group that requires extensive post-secondary education and that is observed 11 years following high school graduation is good evidence that the individuals in the group, on average, are successful and satisfied. That is, mere membership in an occupation under these circumstances is a good criterion. Even the undergraduate major that was the final choice, not the initial one, is a reasonable interim criterion. The graduate major, in turn, represents stronger evidence of a reasonable degree of success and satisfaction. This is especially true among the majors we examine here, inasmuch as degrees in engineering and the physical sciences are among the most conceptually demanding.

Design of the Research

The two abilities in each composite were weighted equally in combined gender distributions in each of the four high school grades. Because of the characteristic gender differences in variances (Lubinski & Dawis, 1992; Stanley, Benbow, Brody, Dauber, & Lupkowski, 1992), the effective weights are only approximately equal for males and females. However, equal weighting was achieved almost perfectly within grades.

The next decision was to select the highest 20% on each selection composite for further study. We then formed three groups defined as follows: High-Intelligence (top 20% on both composites), High-Space (top 20% on Spatial-Mathematics only), and High-Verbal (top 20% on Verbal-Mathematics only). (The top 20% value is more stringent than that imposed for most undergraduate student selection, but engineers and physical scientists are both self- and institutionally selected more highly than other undergraduate majors.)

Sample sizes. The total numbers available for study in the three groups are about equally divided among the four grades but are quite disparate from group to group. For the males, High-Intelligence = 26,908, High-Space = 8,801, and High-Verbal = 7,892; corresponding values for the females were 27,403, 8,846, and 7,972, respectively. The discrepancy in size between the High-Space and High-Verbal groups is the result of differences in size of correlations among the Space, Verbal, and Math components. The much larger numbers in the High-Intelligence group testify to the generality in human intelligence (Humphreys, 1979).

Sampling errors and tests of significance. We have basically two hypotheses:

1. The prediction of group differences is a useful alternative or supplement to the typical regression methodology involving individual differences in predictors and performance criteria.
2. For physical sciences and engineering, the addition of spatial visualization to verbal and mathematical tests adds incremental validity to the prediction of group membership.

We test these hypotheses by looking for patterns of psychologically significant differences between groups that have reasonable sampling stability. On the other hand, this does not require that every statistically significant difference be marked as such or interpreted. We are not alone in concluding that there is no point in computing tests of significance in large samples on sizable differences to conclude that zero differences in something can be rejected at $p < .05$ (Hanisch, 1992; Lykken, 1968; Meehl, 1978).

The problem is somewhat more complex for the follow-up data. Each type of follow-up information is based on different sample sizes in the three subgroups; each sample is also known to be biased on dimensions that characterize those groups. However, sampling weights were available, based on an intensive follow-up of a small sample of nonrespondents conducted by TALENT's staff (Wise et al., 1979), which enabled us to estimate population proportions in educational majors and occupations. Population estimates are reported because they are unbiased even though they lack sampling errors. First, we analyzed unweighted proportions, being guided by the sample sizes that appear in Table 1. For most of the data, it was not difficult to select psychologically important and stable differences for interpretation. Naturally, sample sizes for graduate majors are smallest, but still not tiny. Proportions of graduate majors are also dependent on those for undergraduate majors, which increases the stability of the former when the undergraduate major is known.

Special characteristics of the design. Selecting equal proportions of males and females for study was a matter of convenience. We wanted adequate sample sizes in the 11-year follow-up samples even though the levels of measured abilities were not equal. The use of equal proportions, on the other hand, had nothing to do with affirmative action, assumptions about causation, or ease of remediation, or the psychologi-

Table 1
Sample Sizes for the Four Follow-up Sets of Data, Analyzed by Gender

Category	High-Intelligence		High-Space		High-Verbal	
	Males	Females	Males	Females	Males	Females
Undergraduate major						
N	8,951	7,918	1,888	1,448	2,212	1,799
Graduate major						
N	4,713	3,075	586	390	1,141	630
Occupation						
N	9,393	9,969	2,642	2,836	2,383	2,497
Amount of education						
N	10,584	11,010	2,926	3,086	2,685	2,767

cal importance of the observed gender differences (cf. Benbow, 1988; Lubinski & Benbow, 1992).

Differential effects of spatial and verbal abilities should show up best in groups in which one or the other is not allowed to vary within the full range of talent, *and* mathematical ability is held approximately constant. The High-Intelligence group differs from the other two in having (a) a higher mean on a dimension appropriately called general intelligence, and (b) a flat ability profile. Given this, if an outcome for either of the two contrast groups (viz., High-Space or High-Verbal) is approximately the same as for High-Intelligence, the importance of overall ability for that outcome is diminished.

Results

Basic Information About the Groups

Relations among the variables. The intercorrelations of the selection composites and their components appear in Table 2. These correlations were computed in the total sample of 12th-grade students. Even though mean performance increased substantially during the high school years, the correlations among the variables were almost constant from grade to grade (cf. Lubinski & Humphreys, 1990b, Appendixes A & B). As the correlations in Table 2 reveal, there is indeed generality in human intelligence in a wide range of talent. That is, common psychological processes run through mathematical, spatial, and verbal abilities, but spatial ability has less in common with the other two than they have with each other.

Note also that the correlations between Spatial-Math and Verbal-Math and between a component and its composite are spuriously high. Ideally, the Math scores entering the two composites should be experimentally independent of each other. If independent Math composites parallel to ours were used, and if the correlation between the two were .90, correlations between Spatial–Math and Verbal–Math would be lower by about .03. Thus, our estimates of the proportions of persons who would be found in High-Space and High-Verbal groups are spuriously low and the proportion in the High-Intelligence group spuriously high, but only by modest amounts.

Table 2
Intercorrelations of the Ability Measures for Students in the 12th-Grade Samples

Ability measure	1	2	3	4	5
1. Math	—	.61	.76	.90[a]	.94[a]
2. Space	.62	—	.62	.90[a]	.65
3. Verbal	.79	.60	—	.77	.94[a]
4. Spatial-Math	.90[a]	.90[a]	.77	—	.89[b]
5. Verbal-Math	.95[a]	.65	.95[a]	.88[b]	—

Note. Approximately 40,000 males are below, 40,000 females above, the diagonal; σ_z = .005.
[a] Spuriously high part–whole correlations. [b] These correlations are inflated by the unavoidable use of identical mathematics components in both composites.

Performance levels of the groups. The upper half of Table 3 contains information about the level of performance of our groups relative to the general population of high school students. The distributions of raw scores on these measures, from which the standardized scores were computed, are not completely normal; but this produces no appreciable amount of uncertainty in interpreting the main trends. Note also that gender differences can be obtained by appropriate subtractions.

Data for the 12th graders only are presented again. Keep in mind, however, that the 12th-grade means are not representative of observed means in lower grades because the 12th-grade correlations were of those in lower grades. Means for Grades 11, 10, and 9 would look like those for the 12th grade in 1, 2, and 3 years, respectively (cf. Lubinski & Humphreys, 1990b, Appendixes A & B). Neither the gains nor the gender differences are constant during the high school years, perhaps as a result of patterns of course-taking by year and by gender. The 12th-grade means are representative of the performance of high school seniors for all four high school classes in these data.

The High-Intelligence group, of course, has the highest means on both composites, which is the expected result of the method of selection. As a function of the correlations in Table 2, the three samples selected to be in the upper 20% of either or both of our composites represent, in all, 25% of the high school population. The High-Intelligence group represents, on average, a superior three fifths of that 25%. This group is also quite high on Math (which is a component of both selection composites) but differs little, albeit in different directions, from High-Space (in the Space component) and High-Verbal (in the Verbal component).

Gender differences. For 12th-grade students tested in 1960, there is a very substantial gender difference in Space, one almost as large in Math, and one of trivial size in Verbal. Given

Table 3
Mean Ability Scores of the Three 12th-Grade Subsamples, Standardized in Gender-Combined and Own-Gender Distributions

Ability measure	Males High-Intelligence	Males High-Space	Males High-Verbal	Females High-Intelligence	Females High-Space	Females High-Verbal
			Gender-combined distribution[a]			
Spatial-Math	1.58	1.33	.84	1.03	.70	.19
Verbal-Math	1.47	.77	1.26	1.24	.44	.93
Math	1.78	1.19	1.43	1.29	.46	.76
Space	1.40	1.47	.26	.78	.93	−.36
Verbal	1.16	.37	1.08	1.19	.48	1.10
			Own-gender distribution[b]			
Spatial-Math	1.41	1.14	.61	1.58	1.17	.55
Verbal-Math	1.41	.69	1.19	1.51	.59	1.15
Math	1.45	.91	1.13	1.66	.75	1.08
Space	1.10	1.18	−.04	1.18	1.34	−.07
Verbal	1.17	.40	1.10	1.18	.45	1.09

Note. Any value of .08 for a difference in means by either rows or columns is, very conservatively, significant at $p < .01$.
[a] This metric has a mean of zero and a standard deviation of 1.00. [b] Each of the samples of all 12th-grade males and females has a mean of zero and a standard deviation of 1.00.

the approximately equal weighting of the components in the two composites for the two genders, differences on the composites are approximately equal to the means of the differences on the components. Use of the same qualifying score on either Verbal-Math or Spatial-Math for admission in, for instance, engineering would have an "adverse impact" (in the language of equal employment) on women, but this impact would be substantially greater for the Spatial-Math composite. Even if gender differences were entirely environmentally determined, which is a hypothesis that cannot be rejected by any current data, differences do not quickly and easily disappear after the provision of opportunity at age 18 (Humphreys, 1988).

Means in own-gender metric. It is useful to look at the performance of the two genders when the means of each gender are reported in the own-gender metric. (This metric does not implicitly assume that gender differences in *means* are fixed over time.) These data are in the lower half of Table 3. These means are not far removed from those that one might estimate from the data in the upper half of the table, but differences are produced by the greater variability in all measures of males (cf. Lubinski & Dawis, 1992; Stanley et al., 1992). By and large, these means show that the two genders were selected at about the same levels in their own-gender distributions. The largest differences are associated with the females in the High-Space group, who score a little lower than the males in Math and higher in Space relative to their gender-equivalent normative peers. High-Space females also score a little lower on the Verbal-Math composite, again, relative to the norm for their gender.

Educational and Occupational Outcomes

Data on college majors appear in Table 4 for the three select groups and for the total sample of all high school students (norm) for the four high school grades combined. Note that the norm group contains the other three. The four grades are now combined to have an adequate number in each of the several educational categories. Note that the proportions in the table, furthermore, are population estimates, not sample values.

Undergraduate majors. There are large differences in college majors of High-Space and High-Verbal groups (see the upper half of Table 4). High-Space has more than double for men, more than triple for women, the proportion in the physical science category that includes engineering, mathematics, and computer science. High-Verbal has almost triple the proportion for both genders in the humanities and social sciences. Some extremely interesting differences for the arts also emerge. For young men the proportion in High-Space is double that in High-Verbal groups, and for young women the latter is very nearly quadruple. Is it possible that engineering design has a good deal in common with the arts? When the data for the High-Intelligence Group are considered, it is clear that the common element is not mathematics. Rather, it seems to depend on average scores on verbal tests accompanied by high scores on spatial visualization tests. It is also noteworthy that the High-Space group has proportions in physical science majors that are almost as high as those of High-Intelligence. The latter group has a higher level of general intelligence and is better qualified in mathematics, but higher verbal scores appear to be associated with greater interest in the humanities and the social sciences.

Graduate majors. The lower half of Table 4 contains proportions for graduate major groups defined as they were in the upper half of the table. The male High-Space group is still found more heavily in physical sciences than the male High-Verbal group. All three female groups had small proportions in the physical sciences as a function of role expectations in the

Table 4
Proportions in Four High School Classes of Undergraduate and Graduate Majors of the Three Select Groups and Students in General

Major	Norm		High-Intelligence		High-Space		High-Verbal	
	Males	Females	Males	Females	Males	Females	Males	Females
			Undergraduate majors					
Physical Sciences	.24	.05	.39	.10	.37	.08	.16	.02
Biological Sciences	.11	.12	.11	.13	.10	.11	.12	.12
Business	.25	.09	.16	.04	.23	.08	.20	.03
Education/Social Work	.10	.35	.04	.28	.08	.43	.07	.38
Humanities/Social Sciences	.27	.30	.28	.39	.16	.16	.43	.41
Arts	.03	.08	.02	.07	.05	.13	.02	.03
			Graduate majors					
Physical Sciences	.17	.04	.27	.07	.23	.02	.07	.01
Biological Sciences	.12	.06	.14	.09	.15	.03	.14	.02
Business	.16	.02	.15	.02	.22	.04	.12	.00
Education/Social Work	.20	.63	.10	.52	.16	.67	.18	.69
Humanities/Social Sciences	.32	.21	.31	.26	.16	.17	.46	.26
Arts	.03	.05	.02	.04	.08	.07	.03	.03

Note. These proportions are based on population estimates.

late 1960s and early 1970s. All female groups entered graduate work in education at a high rate, and the proportion of High-Space students was substantially higher than that for High-Intelligence. Both genders in High-Space continued to avoid, relatively speaking, social sciences and humanities. Art majors are, again, attractive to High-Space students, but graduate art majors are few and differences are small.

Occupational outcomes. Proportions in occupational categories appear in Table 5. High-Space compares favorably with High-Verbal in the physical sciences for young men but has lost a little with respect to the High-Intelligence group. The reason is not hard to find. High-Space persons of both genders, but especially men, were working in larger proportion in traditional blue-collar occupations. Engineers, artists, and now artisans have something in common. High-Space women were also working in larger proportions in secretarial-clerical occupations. High-Space persons have a mean at the 88th percentile in mathematical talent and at the 64th percentile verbally, yet they are found disproportionately in occupations for which a high school education is considered sufficient.

Amount of education completed. The proximate cause of the disproportionate numbers of High-Space students in occupations for which a baccalaureate degree is not typically required is shown by the data in Table 6. High-Space has substantially smaller proportions of holders of credentials at every educational level beyond high school graduation in comparison with High-Intelligence and High-Verbal groups. The High-Space group even has a slightly higher proportion of dropouts.

Table 5
Proportions in Four High School Classes of Occupational Categories of Three Select Groups and Students in General

Major	Norm		High-Intelligence		High-Space		High-Verbal	
	Males	Females	Males	Females	Males	Females	Males	Females
Physical Science	.04	.00	.15	.01	.10	.00	.04	.00
Biological Science	.02	.03	.07	.05	.04	.06	.06	.07
Business	.32	.07	.34	.08	.35	.06	.34	.07
Education/Social Work	.07	.08	.09	.18	.09	.10	.14	.19
Humanities/Social Sciences	.04	.01	.11	.04	.02	.01	.19	.04
Arts	.01	.01	.01	.01	.03	.01	.02	.00
Technical	.06	.02	.08	.02	.08	.02	.05	.02
Secretarial/Clerical	.04	.16	.02	.11	.02	.16	.01	.12
Artisan	.39	.11	.14	.05	.28	.07	.12	.05
Housewife	.00	.50	.00	.44	.00	.50	.00	.43

Note. These proportions are based on population estimates.

Table 6
Proportions in Four High School Classes of Amount of Education Completed by the Three Select Groups and Students in General

	Norm		High-Intelligence		High-Space		High-Verbal	
Education	Males	Females	Males	Females	Males	Females	Males	Females
PhD	.03	.00	.11	.02	.02	.00	.08	.01
MA +	.06	.04	.16	.13	.09	.05	.18	.09
BA +	.17	.13	.34	.33	.25	.17	.34	.38
HS +	.53	.64	.26	.41	.50	.65	.27	.41
Dropped out	.05	.06	.01	.01	.01	.04	.01	.01
Uncertain	.02	.01	.03	.02	.01	.01	.02	.02
No response	.14	.12	.10	.08	.11	.09	.10	.08

Note. These proportions are based on population estimates. Ph.D. = Doctoral degree; it includes degrees in law and medicine. The + next to MA (Master's degree in Arts), BA (Baccalaureate degree in Arts), and HS (high school diploma) denotes the inclusion of individuals having course work beyond that level but not enough to achieve the next highest credential.

If more of these persons could be encouraged to work for a college degree, for which they are qualified in terms of measured ability, it seems highly probable that the proportion of art, engineering, and physical sciences majors would be increased disproportionately in comparison with majors in the social sciences, humanities, and preprofessional curricula.

Self-Report Correlates

As we mentioned earlier, students completed a long series of questions about themselves and their families that were used by the Project TALENT staff to form a series of background scores. Means of these scores for 12th-grade students only appear in Table 7 in the own-gender standard score metric. There are, of course, gender differences on these scores, but the size of these differences is more subject to social change than are the scores for ability tests (Lubinski & Humphreys, 1990a). Use of the own-gender metric also allows a more accurate comparison of the experimental groups, which is our primary focus. It also must be kept in mind that the High-Intelligence group, as a function of the design, has a higher level of general intelligence than High-Space and High-Verbal groups.

Family background. High-Space students are from families that are lower in socioeconomic status (SES), in comparison with High-Verbal students, by .24 and .23 standard score units for males and females, respectively. This is about the amount expected from differences in SES correlations between the spatial component and the math and verbal components. Analysis of the items in the SES index reveals that both parents of High-Space students were more likely to be skilled workers and less likely to be proprietors or professionals. Both parents had some-

Table 7
Mean Standardized Scores of the Three Subgroups in the Metric of Own-Gender 12th-Grade Distributions Plus Correlations of Socioeconomic Status (SES) With Other Background Scores

	High-Intelligence		High-Space		High-Verbal		Correlations with SES	
Measure	Males	Females	Males	Females	Males	Females	Males	Females
SES	.67	.65	.32	.27	.56	.50	1.00	1.00
Curriculum[a]	.90	.95	.46	.35	.84	.85	.38	.40
Academic courses	1.08	1.13	.53	.43	.99	.91	.42	.42
Grades	.77	.84	.01	.10	.66	.60	.11	.12
High school guidance	.25	.29	.10	.16	.21	.23	.19	.19
Other guidance	.32	.32	.17	.18	.28	.25	.25	.24
Study habits	.64	.56	.13	.11	.62	.53	.19	.20
Writing skills	.52	.54	−.12	.00	.66	.63	.19	.20
Reading skills	.57	.53	.02	.00	.56	.58	.18	.15
Reading amount	.38	.46	−.02	.04	.37	.49	.15	.18
Variety of hobbies	−.01	.19	.23	.25	−.29	−.07	.04	.14
Activity in hobbies	−.08	.18	.16	.26	−.33	−.03	.03	.16
Work activities	−.23	−.08	.07	.07	−.24	−.14	−.05	−.06

Note. Any value of .08 for a difference in means by either rows or columns is significant at $p < .01$.
[a] Standard scores for this variable are proportions in the precollege curriculum.

what less education as well. Although these differences between High-Space and High-Verbal groups are real, the High-Space group does have above-average SES status somewhat closer to the High-Verbal group than it is to the high school norm. SES is positively related to college entrance, more highly than its merit as a predictor of measured achievement warrants, but the differences just described cannot explain why the High-Space group is so nearly identical with the high school norm in having a high school diploma as its highest educational credential.

On a priori grounds, one might suspect that SES differences would strongly confound interpretations of other differences involving background scores. Yet the correlations of SES with the other scores, found in the rightmost column of Table 7, reject this possibility. Of these correlations, the largest is only moderate and most are trivial in size.

Student biographical scores. Standard score differences between High-Space students and the High-Verbal/High-Intelligence groups are about twice the size of the SES differences for (a) being in the precollege curriculum, (b) the number of solid academic courses taken, (c) level of high school grades, and (d) quality of study habits. These background scores help to explain why members of the High-Space group did not acquire equivalent educational credentials following high school graduation.

There are indications among the items in the measure of academic grades, however, that High-Space students were not generally inadequately motivated to achieve. Their grades in mathematics were about equal to those of High-Verbal students, and their science grades were not far behind. Grades of High-Space students in vocational courses were higher in both genders, and female members had higher grades in business and commercial courses as well. Grades in foreign languages, history and social studies, and English pulled down the grade point averages (GPAs) of High-Space students.

Responses to the hobbies questions throw light on the kinds of people who compose High-Space groups. Members of that group report a greater variety of hobbies and a greater degree of participation in them as well. The differences are smaller for females than for males, but TALENT's questionnaire sampled an overrepresentation of hobbies preferred by men. Differences among the items making up the variety and degree of participation scores are again revealing. Group differences are minimal in acting, singing, dancing, and in team sports as well as in collecting stamps, coins, rocks, and so on. On each of 13 other hobbies listed, however, High-Space males and females report a higher incidence of participation than the other four groups. What these other hobbies have in common can be derived from the verbs used in describing the hobbies, as well as the objects of these verbs. The verbs used include *building, working with, making, repairing, sewing, cooking, drawing, painting,* and *gardening.* The objects are things that are created, shaped, or transformed in some way by the subject. Both genders in High-Space groups worked more in and outside the home than High-Intelligence and High-Verbal groups. Inspection of individual items reveals that more work also resulted in more income and involved every type of work activity listed in the questionnaire.

High-Space students elected to participate in hobbies because they wanted to. Somewhat too many spatially talented students elected the wrong track (or were placed in it) for becoming prime candidates for college admission. This is only a part of the explanation, however, for their deficit in amount of formal education relative to the other groups. The hypothesis that emerges is of students who were "turned off" from formal education by the highly verbal nature of the precollege curriculum. Obviously, this hypothesis requires further study.

Vocational interests. Means for the interest scores are in Table 8. Once again, the means are reported in the own-gender metric. There are large gender differences in scale means, but the own-gender metric provides a clearer picture of what the two genders have in common and of how the experimental groups differ.

The High-Space groups have less interest than the other two groups in biological, public service, and literary-linguistic occupations. Their physical science interests are intermediate with respect to the other groups, but this area remains primary for them. Males in High-Space do not reject the skilled trades and labor occupations as do High-Intelligence and High-Verbal males, and females follow the same pattern for low-status clerical work. We conclude that the High-Space groups have lower levels of occupational aspiration than the rest, which makes their relatively high interest in engineering and physical science the more remarkable. The interests of High-Space students are a part of a broader pattern of lower educational and occupational aspiration and attainment.

Discussion

Considerations for Use of Group Membership and Spatial Tests

Our study documents the utility of the criterion of group membership for evaluating the validity of predictions across a substantial temporal gap. This methodology is to be distinguished from the classic approach to predictive validation, involving the regressions of performance criteria on predictors. Both methodologies can be used in a complementary fashion, however. Future researchers should consider incorporating the ideas of Kelley (1940) and Rulon et al. (1967) into validation designs aimed at predicting group membership. Multiple discriminant function and canonical correlation are two methodologies that provide useful approaches to predictive validation, classification, and selection and also are likely to yield useful information for educational and vocational psychologists.

In connection with the substantive component of our study, we conclude that a measure of spatial visualization administered on a nationwide scale should provide useful, perhaps essential information on students being considered for admission to schools of engineering and several physical science disciplines. Scores on a spatial-visualization composite would probably add incremental validity to verbal and math scores, which are currently being used for identifying students with exceptional talent for engineering and physical science. Moreover, spatially talented individuals not only have the ability to achieve career excellence in engineering and the physical

Table 8
Mean Standardized Interest Scores of the Three Subgroups in the Metric of Own-Gender 12th-Grade Distributions

	High-Intelligence		High-Space		High-Verbal	
Interests	Male	Female	Male	Female	Male	Female
Physical science	.80	.79	.56	.33	.25	.24
Biological science	.41	.52	.10	.16	.33	.39
Public service	.19	.31	−.07	.01	.34	.33
Literary-linguistic	.39	.57	−.15	.03	.54	.63
Social service	.01	.15	−.14	.00	.20	.27
Artistic	.24	.49	.14	.33	.01	.31
Music	.20	.44	−.05	.13	.17	.35
Sports	−.02	.25	−.01	.18	.08	.14
Outdoor recreation	−.16	.23	.09	.25	−.22	.06
Business	−.03	.08	−.06	.05	.12	.13
Sales	−.16	−.04	−.12	.04	.00	.02
Computation	.16	.01	.15	.20	.10	−.28
Office work	−.17	−.56	−.07	−.12	−.16	−.51
Mechanical-technical	−.11	.24	.25	.20	−.53	−.02
Skilled trades	−.41	−.02	−.05	.17	−.55	−.18
Farming	−.19	.27	.03	.29	−.28	.08
Labor	−.41	−.01	−.07	.10	−.45	−.07

Note. Any value of .08 for a difference in means by either rows or columns is significant at $p < .01$.

sciences but they also are more likely to remain committed to these disciplines. Furthermore, although our research was aimed at the more technical sciences, we found that the importance of spatial skills is also seen in many of the creative arts.

Selection on space and mathematics. The prevailing emphasis on verbal scores on national tests and on grades in verbal courses for placing students in the precollege curriculum and in encouraging students to think of themselves as college material might be destructive to those who are intellectually talented in nonverbal ways. Students who are fluent verbally are ideal in the minds of many educational personnel at all levels, and this ideal is readily transmitted to parents and students. The case must be made for another important combination of abilities, and students who are suitably high on that combination should be strongly encouraged to aspire to college training. Consequently, more spatially talented students could be entering technical disciplines (which are highly correspondent to their abilities *and* interests).

In the applied use of a Space-Math composite for admission to relevant curricula, verbal ability could be free to vary and not curtailed at the high end of the distribution, as in our High-Space group. For example, if a highly selective institution recruited the highest 10% of the high school population on a Space-Math composite and ignored the verbal score, those admitted would, in a normal distribution, have a mean standardized score of 1.8 and a percentile of 96 on the selection composite. The mean verbal score (assuming the correlation of .77 reported in Table 2) would be 1.4, which represents the 92nd percentile. Selecting the upper 20% on Spatial-Math would result in standard scores of 1.4 and 1.1 and percentiles of 92 and 86, for the Space-Math composite and the verbal score, respectively.

What are the implications of adopting this alternative selection procedure? Students selected for engineering and physical sciences on the basis of the Verbal-Math composite, which is essentially the SAT and the current model, would certainly obtain higher grades in the humanities and the social and biological sciences than those selected on Space-Math. This advantage would probably extend to highly verbal exams in the physical sciences as well. Large numbers of students with high total scores on the SAT are not interested, however, in engineering and the physical sciences, and many High-Verbal students that do enter these areas quickly transfer to majors more in line with their profile of abilities and interests (Lubinski et al., in press). As a result, the latter disciplines are more heavily dependent on self-selection by students and, consequently, obtain smaller numbers of able students than they would if another definition of talent were accepted and implemented.

Valid nontest information. Although most educational institutions use high school rank or GPA in admissions, we suggest that the pattern of grades (similar to pattern of abilities) makes a difference in assessing a student's appropriateness for training in many disciplines. Schools of engineering, for example, might primarily weight grades in mathematics and science, especially physical science. They might also give substantial weight to grades in vocational courses that require building, repairing, creating, and so on. Participation in hobbies and levels of achievement in hobbies that involve the same activities are also, as we have seen, valid indicators.

The preceding recommendations conform to Standard 9.7 of the test *Standards* (AERA et al., 1985) in that we encourage use of multiple sources of *valid* information and less reliance on standard verbal and quantitative predictors of academic success. Self-report scores are rarely effective predictors of individual differences in educational performance but in our data are effective predictors of group membership 11 to 14 years subsequent to the administration of the tests. We conclude that the latter may be the more important information for counseling, educational, and personnel psychologists.

Possibilities for curriculum change. Our findings also may serve as a source for stimulating innovations in educational curricula. The background data, especially the hobbies data, support the need for more "hands-on" science courses at several educational levels. Are there high school science courses that are offered with minimal laboratory experience? If so, grades in such courses are probably not as meaningful for selecting future scientists, especially physical scientists, compared with courses with challenging laboratories. It also might be useful to prepare a college preparatory course in technology with a large laboratory component. Hands-on contact seems critical for spatially talented individuals. They appear to be attracted to "things" rather than "ideas" or "people" (Prediger, 1976).

Most discussions on the utility of mechanical/spatial tests have tended to be restricted to occupations below the professional level. Based on the findings reported here (coupled with our literature review), we suggest that such instruments be incorporated in selecting students and personnel for high-level technical disciplines at the professional level. Vocational psychologists also should be aware of the relevance of spatial and mechanical abilities when advising their clients about careers in engineering, physical science, and the creative arts. After all, C. P. Snow's (1964) two cultures, which actually mirror Vernon's v:ed and k:m in many ways, are *both* intellectual cultures; it is just that the former is more saturated with verbal/numerical symbols (which applied psychologists are currently assessing admirably), whereas the latter is composed more of content requiring ideation about spatial/mechanical things (which applied psychologists should be assessing and using more admirably).

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1985). *Standards for Educational and Psychological Measurement.* Washington, DC: American Psychological Association.

American Institutes for Research. (1979). *Project TALENT master tape formats.* Palo Alto, CA: Author.

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,* 77–89.

Benbow, C. P. (1988). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature effects and possible causes. *Behavior and Brain Sciences, 11,* 169–183, 217–232.

Carroll, J. B. (1985). Exploratory factor analysis: A tutorial. In D. K. Detterman (Ed.), *Current topics in human intelligence: Vol. 1: Research methodology* (pp. 25–58). Norwood, NJ: Erlbaum.

Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: Harper & Row.

Dawis, R. V., & Lofquist, L. H. (1984). *A psychological theory of work adjustment: An individual differences model and its applications.* Minneapolis: University of Minnesota Press.

Flanagan, J. C., Dailey, J. T., Shaycoft, M. F., Gorham, W. A., Orr, D. B., & Goldberg, I. (1962). *Design for a study of American youth.* Boston: Houghton Mifflin.

Hanisch, K. A. (1992). Play it again. *The Industrial-Organizational Psychologist, 29,* 68–70.

Holland, J. L. (1973). *Making vocational choices: A theory of careers.* Englewood Cliffs, NJ: Prentice-Hall.

Hulin, C. L., Henry, R. A., & Noon, S. L. (1990). Adding a dimension: Time as a factor in the generalizability of predictor relationships. *Psychological Bulletin, 107,* 328–340.

Humphreys, L. G. (1962). The organization of human abilities. *American Psychologist, 17,* 475–483.

Humphreys, L. G. (1979). The construct of general intelligence. *Intelligence, 3,* 369–382.

Humphreys, L. G. (1985). General intelligence: An integration of factor, test, and simplex theory. In B. B. Wolman (Ed.), *Handbook of intelligence: Theories, measurement and application* (pp. 201–224). New York: Wiley.

Humphreys, L. G. (1986). Commentary. *Journal of Vocational Behavior, 29,* 421–437.

Humphreys, L. G. (1988). Trends in levels of academic achievement of blacks and other minorities. *Intelligence, 12,* 231–260.

Humphreys, L. G. (1991). Some unconventional analyses of resemblance coefficients for male and female monozygotic and dizygotic twins. In D. Cicchetti & W. Grove (Eds.), *Thinking clearly about psychology: Essays in honor of Paul Everett Meehl* (pp. 158–187). Minneapolis: University of Minnesota Press.

Humphreys, L. G., Davey, T. C., & Kashima, E. (1986). Experimental measures of cognitive privilege/deprivation and some of their correlates. *Intelligence, 10,* 355–370.

Kelley, T. L. (1940). *Talents and tasks: Their conjunction in a democracy for wholesome living and national defense* (Harvard Education Papers, No. 1). Cambridge, MA: Harvard Graduate School of Education.

Lofquist, L. H., & Dawis, R. V. (1991). *Essentials of person–environment correspondence.* Minneapolis: University of Minnesota Press.

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., & Sanders, C. E. (in press). Reconceptualizing gender differences in achievement among the gifted: An outcome of contrasting attributes for personal fulfillment in the world of work. In K. A. Heller, F. J. Monks, & A. H. Passow (Eds.), *International handbook for research on giftedness and talent.* 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.

Lubinski, D., & Humphreys, L. G. (1990a). A broadly based analysis of mathematical giftedness. *Intelligence, 14,* 327–355.

Lubinski, D., & Humphreys, L. G. (1990b). Assessing spurious "moderator effects": Illustrated substantively with the hypothesized ("synergistic") relation between spatial and mathematical ability. *Psychological Bulletin, 107,* 385–393.

Lunneborg, C. E., & Lunneborg, P. W. (1975). *College major similarity profiles based on selected Washington pre-college measures and vocational interest inventory scores* (Educational Assessment Center Report 76-10). Seattle: University of Washington.

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70,* 151–159.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validation results: The relationship between predictor and criterion domains. *Personnel Psychology, 43,* 335–353.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46,* 806–834.

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

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

Smith, I. M. (1964). *Spatial ability.* London: University of London Press.

Snow, C. P. (1964). *The two cultures: And a second look.* Cambridge, England: Cambridge University Press.

Stanley, J. C., Benbow, C. P., Brody, L. E., Dauber, S., & Lupkowski, A. (1992). Gender differences on eighty-six nationally standardized achievement and aptitude tests. In N. Colangelo, S. G. Assouline, & D. L. Ambroson (Eds.), *Talent development: Proceedings from the 1991 Henry B. and Jocelyn Wallace National Research Symposium on Talent Development* (pp. 42–65). New York: Trillium Press.

Tatsuoka, M. M. (1988). *Multivariate analysis: Techniques for educational and psychological research* (2nd ed.). New York: Macmillian.

Vernon, P. E. (1947). Research on personnel selection in the Royal Navy and the British Army. *American Psychologist, 2,* 35–51.

Vernon, P. E. (1950). *The structure of human abilities.* London: Methuen.

Willerman, L. (1979). *The psychology of individual and group differences.* San Francisco: Freeman.

Wise, L. L., McLaughlin, D. H., & Steel, L. (1979). *The Project TALENT data bank.* Palo Alto, CA: American Institutes for Research.

Received February 24, 1992
Revision received July 6, 1992
Accepted July 8, 1992 ■