

Psychometric
and
Social Issues

Intellectual Talent



Edited by
Camilla Persson Benbow
and
David Lubinski

THE JOHNS HOPKINS
UNIVERSITY PRESS

Baltimore and London

Contents

© 1996 THE JOHNS HOPKINS UNIVERSITY PRESS

All rights reserved. Published 1996
Printed in the United States of America on acid-free paper
05 04 03 02 01 00 99 98 97 96 5 4 3 2 1

THE JOHNS HOPKINS UNIVERSITY PRESS
2715 North Charles Street
Baltimore, Maryland 21218-4319
The Johns Hopkins Press Ltd., London

Library of Congress Cataloging-in-Publication Data will be found at the
end of this book.

A catalog record for this book is available from the British Library.

ISBN 0-8018-5301-X
ISBN 0-8018-5302-8 (pbk.)

List of Contributors	ix
Preface	xi
I. Political Correctness and the <i>Zeitgeist</i> : Genetic Antecedents to Human Behavior	1
1. Genes, Drives, Environment, and Experience: EPD Theory Revised THOMAS J. BOUCHARD, JR., DAVID T. LYKKEN, AUKE TELLEGEN, AND MATTHEW MCGUE	5
2. The IQ Controversy and the Gifted ABRAHAM J. TANNENBAUM	44
II. The Underuse of Knowledge	79
3. Educational Research and Educational Policy: The Strange Case of Acceleration JAMES J. GALLAGHER	83
4. Acceleration over the Years A. HARRY PASSOW	93
5. The Role of the Educational Researcher in Educational Improvement: A Retrospective Analysis HERBERT J. KLAUSMEIER	99
6. Assessing Spatial Visualization: An Underappreciated Ability for Many School and Work Settings LLOYD G. HUMPHREYS AND DAVID LUBINSKI	116
III. What Do We Know about Proper Provisions for the Gifted?	141
7. Motivating Academically Able Youth with Enriched and Accelerated Learning Experiences JOHN F. FELDHUSEN	145
8. Gifted Youth: A Challenge for Science Education LYNN W. GLASS	159

vi	Contents	
9.	Acceleration as an Option for the Highly Gifted Adolescent NANCY M. ROBINSON	169
10.	Acceleration among the Terman Males: Correlates in Midlife and After LEE J. CRONBACH	179
11.	The Elephant in the Classroom: Ability Grouping and the Gifted ELLIS B. PAGE AND TIMOTHY Z. KEITH	192
12.	What Is Learned in School and What Is Learned Outside? JAMES S. COLEMAN	211
IV.	The Use of Knowledge: The SMPY Project	217
13.	<i>Quo Vadis</i> America? ARNOLD E. ROSS	221
14.	In the Beginning: The Study of Mathematically Precocious Youth JULIAN C. STANLEY	225
15.	Contributions of the Talent-Search Concept to Gifted Education JOYCE VAN TASSEL-BASKA	236
16.	Nurturing Exceptional Talent: SET as a Legacy of SMPY LINDA E. BRODY AND CAROL C. BLACKBURN	246
17.	The Impact of SMPY's Educational Programs from the Perspective of the Participant CAMILLA PERSSON BENBOW, DAVID LUBINSKI, AND BABETTE SUCHY	266
18.	Correlates of High Mathematical Ability in a National Sample of Eighth Graders RICHARD E. SNOW AND MICHELE ENNIS	301
V.	Psychometrics: Generality and Specificity	329
19.	The Utility of Out-of-Level Testing for Gifted Seventh and Eighth Graders Using the SAT-M: An Examination of Item Bias CAMILLA PERSSON BENBOW AND LEROY WOLINS	333
20.	Construct Validity of the SAT-M: A Comparative Study of High School Students and Gifted Seventh Graders LOLA L. MINOR AND CAMILLA PERSSON BENBOW	347
21.	The Generalizability of Empirical Research Results BETSY JANE BECKER	362
22.	Possible New Approaches to the Study of Mathematically Precocious Youth N. L. GAGE	384

Contents	vii
VI. Giftedness \neq Genius	391
23. Giftedness and Genius: Crucial Differences ARTHUR R. JENSEN	393
Appendix	413
Index	419

Assessing Spatial Visualization

An Underappreciated Ability for Many School and Work Settings

LLOYD G. HUMPHREYS AND

DAVID LUBINSKI

A Brief History of Spatial-Ability Testing

Spatial Visualization in Intelligence Tests

The assessment of individual differences in spatial visualization began concomitantly with the building of early measures of general intelligence. The first Binet-Simon test (1905) contained items that today would be identified as primarily spatial in content, as did the first Stanford-Binet test (Terman, 1916). Back then, item selection was based mostly on content validity, but an empirical criterion was also consulted before an item was accepted—namely, the percentage passing with age. By the time the 1937 revision of the Stanford-Binet test appeared (Terman & Merrill, 1937), more-sophisticated techniques for item inclusion were available. Item selection depended on the percentage passing with age and on correlations between items and the total score on the test, and spatial items were still retained.

McNemar (1942) obtained item intercorrelations and computed centroid factor loadings of the items to define more clearly the number of dimensions necessary to capture their commonality. Across the many different kinds of content found in this next generation of intelligence tests, all item intercorrelations were positive. The first centroid factor, therefore, was large. Yet factors

beyond the first were somewhat larger than expected for Spearman's (1904, 1914) single-factor solution, and the second and third centroids were generally interpretable. Because the item intercorrelations computed in the early days of test construction were typically tetrachorics (correlations based on dichotomized variables), considerable "noise" was introduced into the factor analyses by the variability in size of sampling errors as a function of variability in the percentage passing the items. Nevertheless, in McNemar's (1942) analyses, there was ample evidence to support the construct of general intelligence plus limited evidence for group factors, including spatial visualization.

Around the same time, Wechsler (1941) introduced an intelligence test that provided separate verbal and performance IQs as well as a total IQ. Spatial visualization contributed substantially to the variance of his performance items. This was true also of group tests of intelligence that provided both verbal and performance scores (Vernon, 1947). Moreover, regardless of the test, these separate verbal and performance IQs were not correlated with each other nearly so highly as their reliabilities would allow. Across a wide range of intellectual talent, the correlation is and always has been substantial (r s tend to be around .70 to .80). Although this supports the construct of general intelligence, neither the verbal nor the performance score alone is an adequate measure of general intelligence. In addition, the profiles of correlations of the two scores with other tests and important social criteria have not been identical. These two measures and the constructs they assess have differential validity across many different criteria commonly valued in educational and vocational contexts (Humphreys, 1962, 1986; Lubinski & Dawis, 1992). Thus performance tests of intelligence are only partial substitutes for a standard test when dealing with a person having limited proficiency in the language of the tests.

Spatial-Ability Testing in the Military

The Army Alpha and Beta tests of World War I were designed to measure general intelligence, yet they had the same limitations as Wechsler's verbal and performance IQs for assessing it. Alpha was designed for persons literate in English; Beta was designed for those who were not. Beta contained a number of spatial items, but Alpha contained few; Beta's total variance overlapped only partially with the variance of Alpha. Again, neither test alone was a comprehensive measure of general intelligence in the Binet tradition. However, as with Wechsler's (1941) verbal and performance IQs, aggregating these two early measures does provide a respectable index of general intelligence.

By World War II, a spatial-visualization test was in use by the military. It was included in the Army General Classification Test (AGCT). The ver-

bal, quantitative, and spatial items were printed and administered in a quasi-random fashion and a total score was obtained. In this and similar tests, it was common for different examinees to receive the same total score based on dissimilar profiles of success in the three areas. This is not a problem with respect to the interpretation of the total score as an estimate of general intelligence. It becomes a problem, however, when quantitative-, spatial-, and verbal-ability markers of general intelligence are not given differential weights in predicting performance criteria (or group membership) for which they have differential validity (e.g., technical versus clerical occupations). This became apparent in the selection and classification of air-crew candidates. Here the military developed a special qualifying test (not designed to measure general intelligence) and then followed it with a multiple "aptitude" test battery.¹ Tests in the latter were weighted on the basis of multiple regression analyses, and separate composites were formed for each different air-crew assignment. Tests that received zero or positive weights for pilot selection, a spatially saturated occupation, are highly informative with respect to the aims of this chapter. Reading comprehension, arithmetic reasoning, and mathematics received zero weights. The positive weights for pilots were on tests measuring visual perceptual speed, spatial visualization, spatial orientation, mechanical information and comprehension, large-muscle coordination, and information about planes, flying, and pilots.

The information test, in its final form, was especially revealing psychologically, because it included a few questions involving literary and artistic information, the *wrong* answers to which were given the same positive weight as information about the P-51 airplane. By *suppressing* the variance of these specialized verbal items, the correlation between the total score on the test and performance in the air was increased. If the criterion had been performance in ground school, however, this effect, as well as the test weights, would have been reversed. That is, the suppressor emerged in the context of forecasting terminal criterion performance, but not in the context of forecasting achievement criteria that were antecedents to the terminal performance.

An experimental group that entered training without attention being paid to their test scores had a mean AGCT score of 113 and a standard deviation of 14 (in a metric with $M = 100$ and $SD = 20$). Thus, self-selection was responsible

1. The aptitude/achievement distinction is one of psychology's conceptual ghosts (Cleary, Humphreys, Kendrick, and Wesman, 1975). It turns out that what distinguishes "achievement" tests from "aptitude" tests is a difference in degree, along four dimensions, and not a difference in kind. In contrast to achievement tests, aptitude tests typically: sample from a broader range of content, are not tied to a specific educational curriculum, sample old learning, and are used for forecasting future performance as opposed to determining current status with respect to concurrent criterion behavior. Indeed, items from achievement and aptitude tests are often used interchangeably.

for a mean more than three-fifths of a standard deviation above the mean for the population at large. Those who also passed the qualifying test had a mean AGCT score almost one standard deviation above the overall mean. On the completion of training, the pilots were commissioned as second lieutenants. They were professionals in terms of their credentials and in terms of their levels of general intelligence, but they were *especially* gifted in spatial-visualization and mechanical-reasoning abilities. The importance of their professional status will become clearer as our discussion unfolds.

Post-World War II Military Use of Spatial-Ability Testing

The postwar Armed Forces Qualifying Test (AFQT) continued to involve the tripartite content of verbal, quantitative, and spatial items until the early 1950s, when a fourth item type, mechanical information, was introduced. One of us (Humphreys) took the lead in urging colleagues in the other services to make that addition. The demand for personnel was greater in mechanical-technical assignments than in the more academic ones, because the former were generally more critical to the military mission. Even though mechanical-information tests were already prominent and well validated in the so-called aptitude indexes used in classifying enlisted personnel, the addition of information items to an intelligence test was not accepted enthusiastically by military psychologists. Tradition required that the initial selection of military personnel be based on "aptitude." After several years of increasing somewhat the proportion of enlisted personnel who qualified for the most urgent and important assignments, mechanical information was removed from the AFQT. But it was not removed from the mechanical composite (the aptitude index), which guided the assignment of personnel to the relevant occupational specialties.

Subsequently, the spatial-visualization section of the AFQT was removed. It was also removed from the Armed Forces Aptitude Battery at the same time. Both the mechanical-information and the spatial-visualization items have an adverse impact on the scores of female applicants for military service, but perhaps lower means were more easily tolerated when they occurred on information tests than on "aptitude" tests. In any event, both these decisions were wrongheaded. Women whose scores are high with regard to their same-sex norms on mechanical and spatial tests are the ones needed by the military services to fill the increasing number of specialties open to women. At times, even psychologists forget that testing is an empirically grounded technology.

Spatial-Ability Testing in Industrial Selection

Spatial tests appeared early in the repertoire of tests used by industrial psychologists. A prominent source of training in industrial psychology and of ideas

about what industrial psychologists should do was Donald G. Paterson of the University of Minnesota. Paterson also spearheaded the development of the Minnesota Paper Form Board, a well-known spatial test that is still in use (Paterson, Elliott, Anderson, & Toops, 1930). Some of the psychomotor tests used by industrial psychologists also had spatial content. Paterson, a giant in the early vocational-guidance movement (Paterson, 1938), imbued students with respect for data, and Minnesota became known as the home of “dust bowl empiricism,” a term popularized by psychologists whose conception of theory was more humanistic than scientific.

The Accepted Role of Spatial Tests

Unfortunately, spatial tests became stereotyped as suitable only for personnel selection and vocational counseling in connection with occupations below the professional level. Studies of their validity were almost entirely restricted to assignments for enlisted personnel in the military and skilled technological-mechanical jobs in industry. There are few courses offered in high school, and fewer still in college, for which spatial tests will predict grades as well as verbal and quantitative tests do. Both the Scholastic Aptitude Test and the American College Test predict college grades reasonably well in light of the amount of “noise” one can expect to find in college grade point averages. These tests achieve such prediction success without spatial content. This might give the impression that there is no reason to consider spatial abilities when pursuing careers that require college credentials (but this will be refuted below).

Spatial Visualization in Factor Analysis

Spatial-visualization abilities secured a prominent place in the military and in applied-psychology circles in civilian life during both world wars; this also was true for basic psychological science. Spearman (1904) was responsible for the mathematical theory of general intelligence; later, Burt (1940) added group factors to Spearman's g . In the factor-analytic tradition in Britain, for which Spearman and Burt were largely responsible, there was no doubt about the importance of general intelligence. Nevertheless, the British recognized group factors such as verbal, numerical, and spatial abilities before Thurstone's development of centroid analysis. American users of the centroid method tended to operate at a different level of analysis, and thus were able to ignore the general factor by rotating their factors to orthogonal simple structure (that is, each variable was primarily related to only one factor, and the factors were relatively independent). American factor solutions, however, generated tests having positive intercorrelations because they possessed common general-

factor variance. They were also each saturated with items whose content and nature (figures, numbers, or words) readily engendered distinct labels. Indeed, Thurstone's “Primary Mental Abilities” (1938) presented a spatial factor among the nine that he interpreted and among the seven about which he felt most confident. In addition, there were tests with spatial content waiting in the wings, in the form of one of the three additional factors that were rotated but not interpreted.

Soon thereafter, however, Thurstone started recommending and using oblique rotations. This step allowed him to factor correlations in two or more orders. It remained for Schmid and Leiman (1957) to develop a methodology that reconciled a general factor on which tests had factor loadings (the British tradition) with a general factor that hitherto could only be inferred from factor intercorrelations (the Thurstone heritage). The Schmid-Leiman methodology leads inevitably to a hierarchical model of intelligence, having a single general factor at the top of the hierarchy as long as the R -matrix is positive. Schmid and Leiman's (1957) contribution, when used on a large R -matrix, provides support for Vernon's (1950) hierarchical model. This model takes the best of what both the British and the American traditions have to offer. It recognizes the British emphasis on general intelligence (associated principally with Spearman and Burt), and the American emphasis on multiple group factors (associated principally with Thurstone).

Vernon's model provides for major group factors immediately below the general factor. The major factors break down, in turn, into minor group factors. Two major group factors in Vernon's model are verbal-numerical-educational ($v:ed$) and mechanical-spatial-practical ($k:m$). The tests used for college admission in this country measure the general factor, in part, plus the first major group factor. In contrast, the tests weighted positively for the selection of pilots in World War II measured the general factor, in part, but substituted Vernon's second major factor, $k:m$, for the first, $v:ed$. At generally lower levels of both scores and prestige, tests of the first major factor, $v:ed$, predict success in clerical assignments somewhat more accurately than does a test of general intelligence, while tests of the second major factor, $k:m$, have the same pattern for mechanical-technical assignments (Humphreys, 1986; Thorndike, 1994).

Sources of Interest in Spatial Abilities

Military experience. The manifest importance of spatial abilities was revealed in the Aviation Psychology Research Program of World War II and in postwar research on assignments of enlisted military personnel. Spatial visual-

ization and spatial orientation were clearly distinguishable, although they just as clearly belonged to the same family, in research on air-crew assignments. A ranking of the importance of assignments of enlisted personnel to the military mission relates quite accurately to the ratio of the predictive validities of spatial to verbal tests.

It may seem strange that spatial tests given in the military are better at predicting success in military training than are such tests given in high school. The probable explanation is the availability of current military hardware for training purposes, more hands-on experience, less dependence on textbooks, and generally a greater motivation to succeed on the part of able persons who were not motivated by the high school curriculum. Military experience suggests the possibility that civilian secondary and college education is not sufficiently supportive of a society heavily dependent on technology. Is our educational system geared to the production of clerks rather than mechanics? Are we overlooking persons who are talented on Vernon's spatial ($k:m$) major group factor? Have we paid too much attention to an utterly false dictum concerning verbal abilities that consigns many highly able persons to second-class status in the intellectual hierarchy of occupations? The following statement is often repeated: "If students can't write, they can't think." We believe this is nonsense.

The criterion to be predicted. One of us (Humphreys) has been interested for more than forty years in being able to predict membership in criterion groups. Professor Philip J. Rulon was fond of saying that a guidance counselor who followed the logic of multiple regression faithfully would never advise a student to consider engineering. The reason, of course, is that the chances of success as a technician are higher than those as an engineer at any score level on the predictor. The alternative to multiple regression is the multiple discriminant function. The guidance counselor suggests to the examinee that the occupation indicated is the one for which the examinee is close to the centroid of successful, satisfied members of an occupational group. Vocational psychologists also should be interested in such persons and in those who have the highest probability of short-term success according to conventional regression forecasts.

It is reasonable to characterize regressions of proficiency measures on score distributions of predictor tests as snapshots taken at particular points in time. A recent review of the stability of criterion performance over time (Hulin, Henry, & Noon, 1990) provides a rationale for this characterization. In contrast, group membership is a truly *cumulative* criterion. A professional engineer, for example, has survived numerous institutional decisions, starting with

grades in high school courses, graduation, college entrance, college graduation, being hired, and being promoted. Of equal importance are the series of personal decisions about course selection, choice of college, choice of major, persistence in pursuing a degree, and staying in engineering once on the job. We were interested in examining educational and vocational tracks that might require exceptional amounts of spatial-visualization talent. This led to the research that follows.

We describe some recent research related to individual differences in spatial measures and, in turn, how these measures relate to group-membership criteria. These data also contrast other tests, such as mathematical and verbal measures, with the spatial tests in the prediction of group membership. All data were obtained from the Project Talent Data Bank (Flanagan et al., 1962; Wise, McLaughlin, & Steel, 1979).

All three of the following studies involve prediction of group membership as reported in Project Talent's eleven-year follow-up after high school graduation. The length of time following administration of the predictor tests varies from eleven to fourteen years as a function of the examinees' grade in school in 1960 (grades nine through twelve). The studies will reveal that the group-membership criterion is an important one for documenting the validity of predictor tests. Group-membership data complement conventional criteria such as individual differences in criterion performance. We offer this methodology here to supplement, not to supplant, conventional test-validation methods.

First, however, let us respond to a potential concern. Our longitudinal data were collected more than twenty years ago. It is reflexive on the part of many to dismiss the use of old data for any purpose, but a distinction between mean levels of performance and correlations (structural relations among variables) is essential. Means of psychological tests do change over time, but correlations are relatively resistant to cultural change and to cultural differences. Our interest here and in subsequent research was structural relations among variables. The problem of changes in means was met with the use of same-sex standardization scores.

Data

Self-Selection on the Spatial Dimension

A serendipitous discovery by Humphreys, Davey, and Kashima (1986) is relevant here. These authors used the extensive student information in Project Talent's tenth-grade sample to develop scoring keys to measure the construct of

intellectual privilege/deprivation (P/D), a measure designed to be a salient covariant of traditional socioeconomic status (SES) measures, but whose content was restricted to environmental features thought to be especially conducive to intellectual development. In addition, these investigators formed composites of ability tests to measure Vernon's two major group factors and the general factor. Although the communality of *v:ed* and *k:m* are included in the construct of general intelligence, the composition of each of the three measures was experimentally independent of each of the others. These three composites were the criteria against which items of biographical information were validated and keyed to form three possible P/D scales.

Results on the privilege/deprivation (P/D) scales. Keys were formed for males and females separately in male and female subsamples and cross-validated in independent subsamples. The correlations of the four general-intelligence keys with the criterion of general intelligence were homogeneous about a median value of .63, which was substantially higher than the median correlation of intelligence with the measure of socioeconomic status of the student's family (.41). Thus, there is more information about family background that is associated with children's general intelligence than is available in a standard measure of SES.

There is only a little evidence to suggest that a privileged background for intelligence differs from privileged backgrounds for the major group factors. There was only a bit of differentiation between the general-intelligence key and the verbal-numerical-educational P/D key. The differentiation was so small that the authors dropped the latter P/D key from further analysis. A small amount of marginally dependable differentiation was obtained, however, for the general-intelligence P/D key and the mechanical-spatial-practical P/D key.

Results supporting self-selection. In the search for correlates of the P/D keys, data from the eleven-year follow-up also were used by Humphreys, Davey, and Kashima (1986). These data shed only a little light on the P/D keys, but postsecondary data concerning college education and occupation revealed surprising information about the ability composites. We discuss these findings in conjunction with some new data that we present here.

Humphreys, Davey, and Kashima (1986) reported mean standard scores in the metric of the follow-up sample for male and female occupational groups in the physical sciences. We supplement these earlier results here by presenting data on undergraduate majors (table 6.1). Keep in mind that these standardized

TABLE 6.1. *Self-Selection on a Mechanical-Spatial Dimension Indicated by Means in the Standardized Metric of Unselected Tenth-Grade Students*

	General Intelligence		Mechanical-Spatial	
	Males	Females	Males	Females
Undergraduate majors				
Physics ^a	1.17	— ^b	1.28	— ^b
Engineering	.89	— ^b	.94	— ^b
All physical sciences	.88	1.05	.86	1.15
Occupations				
All physical sciences ^a	1.04	1.30	.97	1.16

^aIncludes mathematics and computer science.

^bSample sizes are too small for meaningful results.

scores are in the same-sex metric and were based on the entire tenth-grade sample of boys and girls.

There is relatively little difference between mean levels for both sexes on the intelligence and mechanical-spatial-practical composites in undergraduate majors and occupational groups. Moreover, there is little difference when all physical sciences are considered for both males and females. In the data for males for physics and engineering majors, however, the means for the mechanical-spatial composite are higher than the means for the intelligence composite. This points to self-selection into occupations or fields of study as a function of ability.

The argument for self-selection into the physical sciences and engineering on the spatial dimension is straightforward. The measure of general intelligence has standard content: two parts reading comprehension, one part arithmetic reasoning, and one part abstract figural reasoning. Means for general intelligence increase from high school to college and from college entrance to college graduation as functions of educational curriculum and institutional selection. The correlations of the intelligence and mechanical-practical-spatial composites are .63 and .67 in the male and female samples, respectively. If selection based on spatial ability were only incidental to selection based on general intelligence, the expected mean for the former would be no more than two-thirds the size of the intelligence mean. Yet this was not observed. Why? Explicit institutional selection would have required spatial variance in the high school grades and entrance tests used in college admission to explain the virtual equality of the means, but spatial components in these measures were lacking. Thus, this cannot explain the equality of means. That students were self-selecting engineering and the physical sciences based on their spatial-visualization abilities is most probable. This was the general conclusion of Humphreys, Davey, and Kashima (1986), which is supported by the data presented here.

A Comparison of Two Possible Selection Composites

The next research project followed up on the above serendipitous findings, and is reported more fully in Humphreys, Lubinski, and Yao (1993). Our objective here was to obtain correlates of two possible selection composites—a traditional verbal-quantitative measure and a second measure containing spatial-visualization and mechanical-reasoning variance. The latter, we suspected, would be optimal for the selection of most engineers and physical scientists. The criterion selected for the comparison, group membership, was the same one as in the serendipitous findings.

Methods and rationale. In Humphreys, Lubinski, and Yao (1993) we formed three composites from multiple short tests; each test in its own way measured verbal, mathematical, and spatial abilities or group factors. This was done separately for each sex and for each of the four high school grades. Two selection composites were then formed in each sex and in each grade from equally weighted verbal and mathematical (verbal-math) and spatial and mathematical (space-math) components. Students in the top 20 percent on either the verbal-math or the space-math composite, or both, were selected for further study. Approximately twenty-seven thousand of each sex in all grades were in the upper 20 percent on both composites, while approximately nine thousand were in the upper 20 percent on space-math only and eight thousand were at the same level on verbal-math only.

We realize that the space-math composite does not conform to the description of Vernon's second major group factor, but it seemed inconceivable to consider a selection composite for engineering that did not have a mathematical component. It was our working hypothesis that mathematical abilities are the most important abilities for securing educational credentials in engineering and the physical sciences, but that spatial abilities are also critical, and are more important than verbal abilities. In this case, the math component could not be measured with experimentally independent components. Thus, the overlap between space-math and verbal-math scores is spuriously high. Independent components would provide somewhat greater differentiation, but the potential amount is not large because all the components have substantial loadings on the general factor. The somewhat inflated correlations between the two composites are .88 in the male sample and .89 in the female sample.

Scores on the two composites for the three groups are presented in table 6.2. The upper half of the table contains data in the joint-sex metric, showing the difference in means for the two sexes that existed in 1960. The data in the

TABLE 6.2. Mean Ability Scores of Twelfth Graders in Three Ability Groups, Standardized in Joint-Sex and Same-Sex Distributions

Composites	Male Groups			Female Groups		
	High Intelligence	High Space	High Verbal	High Intelligence	High Space	High Verbal
Joint-sex						
Spatial-math	1.58	1.33	.84	1.03	.70	.19
Verbal-math	1.47	.77	1.26	1.24	.44	.93
Same-sex						
Spatial-math	1.41	1.14	.61	1.58	1.17	.55
Verbal-math	1.41	.69	1.19	1.51	.59	1.15

lower half of the table are in the same-sex metric. This information is presented to undermine explicitly any assumption that means are stable over time.

Our criteria for the selection of samples resulted in approximately the same degree of superiority within sex for both boys and girls. The high intelligence group is also shown to be higher in space-math than the high space group and in verbal-math than the high verbal group. Persons in the upper 20 percent on both verbal-math and space-math are highest in general intelligence and would be expected to have many educational and vocational options open to them.

Results. Table 6.3 contains data on the largest differences in undergraduate and graduate majors for three groups of high school students: a high general-intelligence group who were in the top 20 percent on both composites, a high space-math group, and a high verbal-math group. The omitted major groups showed little effect of group membership.

For those majoring in the physical sciences, including engineering, there is little difference between the high intelligence and the high space-math group in their relatively high proportions, but there are large differences between the latter groups and the high verbal-math group. For those majoring in the humanities and social sciences, the high space-math group is low, the high intelligence group is intermediate, and the high verbal-math group is high. The principal determinants among abilities for choice of a major were the differences in levels of spatial and verbal abilities. Skill in mathematics alone does not incline students toward engineering.

Table 6.4 contains data on membership in occupational groups and on the highest educational credential earned for each of the three groups. The proportion of space-math in occupations in engineering and the physical sciences

TABLE 6.3. *Proportions in Four High School Classes of Undergraduate and Graduate Majors of Three Ability Groups*

	High Intelligence		High Space		High Verbal	
	Males	Females	Males	Females	Males	Females
Undergraduate majors						
Physical sciences, engineering	39	10	37	08	16	02
Humanities, social sciences	28	39	16	16	43	41
Arts	02	07	05	13	02	03
Graduate majors						
Physical sciences, engineering	27	07	23	02	07	01
Humanities, social sciences	31	26	16	17	46	26
Arts	02	04	08	07	03	03

Note: Decimal points omitted.

TABLE 6.4. *Proportions in Four High School Classes of Occupational Categories and Highest Educational Credentials of Three Ability Groups*

	High Intelligence		High Space		High Verbal	
	Males	Females	Males	Females	Males	Females
Occupational category						
Physical sciences, engineering	15	01	10	00	04	00
Humanities, social sciences	11	04	02	01	19	04
Arts	01	01	03	01	02	00
Artisans	14	05	28	07	12	05
Highest education credential awarded						
Ph.D.	11	02	02	00	08	01
M.A.	16	13	09	05	18	09
B.A.	34	33	25	17	34	38
H.S.	26	41	50	65	27	41

Note: Decimal points omitted.

drops relative to high intelligence, but is still well above that of verbal-math. Science-math is high, however, in the artisan category, and the explanation appears in the educational credentials earned beyond high school graduation. The space-math students had levels of credentials lower than those of the other groups.

This prompted us to go back to data obtained when the subjects were in high school to discover the kinds of persons the space-math students were. Their self-reports of grades in mathematics were similar to those in the high intelligence group, in the sciences only a little lower, but much lower in foreign languages, English, and social studies. On the other hand, their grades in vocational courses were the *highest* of the three groups. High space-math students

were less likely to be in the college preparatory curriculum, had parents of somewhat lower social status, and participated more actively in hobbies. Even among the hobbies listed, there are differences. High space-math students selected hobbies that involved building and working with things: sewing, cooking, drawing, painting, and gardening. In Prediger's (1976) map of the world of work, spatially talented people tended to coalesce around the "things" sector of his fourfold Data-People-Ideas-Things model. At this point, we wondered if these observations would generalize to more-advanced educational levels, such as those found in graduate training.

Securing More-Advanced Educational Credentials: Graduate Majors

The above results encouraged further study of the possibilities of predicting group membership (Humphreys & Yao, unpublished manuscript). This time graduate majors were selected as the criteria because of the national interest in sources of scientific and engineering talent. This concern is primarily directed at the physical sciences, including engineering, and arises from several sources: the ratio of American citizens to total enrollees in these areas, and the continuously relatively low enrollment therein of majority females and of blacks and Hispanics of both sexes (see *Science*, volumes 258 [13 November 1992]; 260 [16 April 1993]; 262 [12 November 1993]; 263 [11 March 1994]).

Methods and rationale. From Project Talent, eight groups of male graduate majors were selected from the information provided by the follow-up conducted eleven years after the high school graduation of the four high school grades tested in 1960. These graduate majors were divided into the following groups: the physical sciences, the biological sciences (including medicine), the social sciences, law, engineering, humanities, education, and business. There were fewer female graduate majors, necessitating the dropping of the law and engineering female groups, and adding a small number of undergraduate majors in the physical and biological sciences to the female graduate groups to increase sample size. These latter additions did not appreciably decrease the mean level of these groups below the level of the males on the general factor. An additional group of females who had recorded an undergraduate major in business did reduce the general ability level for the business graduate major group, because these additions inadvertently included "majors" in postsecondary commercial training.

The selected predictors of graduate group membership were the individual cognitive-ability test scores, excluding composites, and the self-reported

personality, interest, and background scores in Project Talent's Data Bank. The CANCOR (canonical correlation) program of SAS was used in each of the four grades for the cognitive and self-report scores separately in all but one analysis. For the ninth graders, the two sets of predictors were combined in a single discriminant analysis in order to determine whether the cognitive predictors would add accuracy to the level obtained by the self-reports alone. As it turned out, it was beneficial that most of the analyses were conducted on the separate sets of predictors. The canonical functions were more readily identified psychologically in the separate sets, presumably because of the low level of cross-correlations between sets that made almost identical discriminations among major groups. (The low cross-correlations attenuated the correlations of all predictors with the canonical functions.)

Results. The variance accounted for among males and females by four functions is presented for each in table 6.5. It can be seen that the cognitive variables accounted for more variance among males than among females. Over and beyond the larger number of male groups, boys in 1960 had wider course selection and vocational choices than girls.

The greater predictive accuracy of the self-report scores as compared to the cognitive predictors has two sources. Across all graduate major groups, both male and female, the mean level of intelligence in the ninth grade is about one standard deviation above the grade mean, and the variance is about half the variance at the same grade level. Yet the variance of the self-report scores is about as likely to be increased as it is to be decreased by the selection on the intelligence dimension. The first source, therefore, is the attenuation of predictive validities of the cognitive tests. The second source is the increment to predictive validity of the self-report measures arising from the cognitive restriction in range, which places an ability floor under the samples. Interest scores in the full range of talent are secondary or tertiary to ability in importance be-

TABLE 6.5. *Variance Accounted for by Four Canonical Functions in Cognitive, Self-Report, and Combined Predictions*

Grade	Cognitive		Self-Report		Combined	
	Males	Females	Males	Females	Males	Females
9	.39	.32	.42	.39	.56	.51
10	.40	.36	.48	.46	— ^a	— ^a
11	.46	.38	.52	.51	— ^a	— ^a
12	.47	.42	.60	.60	— ^a	— ^a

^aData not obtained.

TABLE 6.6. *Key Cognitive Predictors Defining the First Two Functions and the Principal Major Groups Contrasted*

First Canonical Function		Second Canonical Function	
Males	Females	Males	Females
Electronics	Literature	Literature	Electronics
Mechanical reasoning	Social studies	Social studies	Mechanical reasoning
Visualization-3D	Physical sciences	Bible information	Visualization-3D
Introductory math	Introductory math	Vocabulary	Introductory math
Advanced math	Advanced math	Reading comprehension	Mechanical information
Engineering, Physical sciences	All but education	Humanities, law	Physical sciences
versus	versus	versus	versus
Social sciences, humanities, education	Business	Education	Social sciences, humanities

cause so many low-ability students express interest in intellectually demanding occupations.

The increasing maturity of interests and the opportunity to select courses as students move from the ninth grade to the twelfth grade increase the validity of both sets of predictors to discriminate among group means. By the twelfth grade, an inference that the amount of variance accounted for by the combined sets of predictors would be somewhat larger than the .60 values for the self-report scores standing alone seems reasonable. Even if the gain in accuracy were smaller in the twelfth grade than in the ninth grade, this could indicate the effect of differences in cognitive profiles on the development of interests.

Table 6.6 identifies the first two discriminants among the cognitive variables. These are the data most relevant to the topic of this chapter. The first canonical function for males is defined by a combination of mathematical, spatial-visualization, and technical-information tests; verbal tests have only moderate correlations with this dimension. The first canonical function for females, in contrast, is defined primarily by verbal and mathematics tests; spatial visualization has modest negative correlations with the function. (The first component in each sex picks up the variance associated with the differences in general intelligence in addition to other differential variance.) The first female canonical function represents the ideal combination of scores for college entrance as defined by the Scholastic Aptitude Tests, but the first male canonical function is quite different.

The means of the major groups contribute to the interpretation of the first two functions. For males, engineering and the physical sciences are opposed to

the social sciences, humanities, and education, with law, the biological sciences, and business being in the middle. For females, the four academic disciplines are contrasted with business, with education being intermediate.

The second male canonical function is similar to the first for females, although it appears to be even more highly verbal in content. Mathematics tests have correlations with the function of intermediate size, while those for spatial tests are close to zero. The loading of socioeconomic status is larger on the second function than on the first, even though the first function contains more variance in the general factor. Humanities and law are contrasted with education on this function.

The second female canonical function contributes less variance to differentiation, but it is still highly similar to the first function for males. Those majoring in the physical and biological sciences are contrasted with those majoring in the social sciences and humanities, with education and business majors being approximately intermediate. For females, the four academic disciplines are contrasted with business, with education being intermediate.

Discussion

Well-Supported Conclusions

The importance of spatial visualization. Spatial visualization is a more valid predictor of group membership than one would expect from its history in predicting course grades in college preparatory courses in high school, college grades in the physical sciences and engineering, and success in technical training in the military. This does not mean that spatial visualization does not show differential validity in the regression sense in several areas of such technical training. It does, but the amount is small (McHenry, Hough, Toquam, Hanson, & Ashworth, 1990). When used in civilian occupations, however, the sample size is typically too small to show a small gain. Hunter (1983) speculated that performance predictions in scientific-technical disciplines are the ones most likely to profit from spatial-visualization assessments (over and above the general factor). Our findings were consistent with this view. In our comparison of two selection composites, we showed clearly that mathematics is not sufficient for distinguishing between future engineers and future humanists. A high level of mathematics in the presence of high verbal ability and lower spatial ability led to the selection of careers in the humanities and the social sciences, while high mathematical and spatial ability combined with lower verbal ability led to highly technical scientific careers. In the prediction of male graduate majors in

science and engineering, for example, the first canonical function is equally weighted by mathematics and spatial-technical-mechanical tests. Verbal ability has a positive weight, but at about the level one would expect from shared variance on the general factor of intelligence. The same pattern appears on the second canonical function for females when predicting graduate majors in the sciences. Its lower contribution to variance for females compared to males seemingly reflects the lack of attraction to the physical sciences and engineering in the occupational plans of girls in the 1960s. This is true today as well (Lubinski, Benbow, & Sanders, 1993).

Sex differences in abilities. There were large sex differences in mathematics in 1960, which were larger yet in the spatial-technical-mechanical cluster. Sex differences in means are not engraved in stone, so we consistently used same-sex norms in our analysis to allow for changes in mean levels. A good deal of not completely adequate evidence indicates that mean changes have occurred (Feingold, 1988), and there has been change in the choice of college majors and in occupational aspirations (Lubinski, Benbow, & Sanders, 1993). Nevertheless, high school girls still score lower than boys on the cluster of tests involving the physical sciences, the biological sciences, and engineering (Stanley, Benbow, Brody, Dauber, & Lupkowski, 1992).

Girls at the upper end of their same-sex distributions in mathematical-reasoning and spatial abilities are better bets for physical science and engineering majors and occupations than those with high verbal scores in joint-sex distributions, regardless of where the male and female means are at any given time. The profile of the girls' scores predicts choice and a reasonable degree of success and satisfaction. A measure of an important function should never be discarded simply because it shows a mean difference between demographic groups (Humphreys, 1988, 1991a).

Cognitive versus self-report predictors. In groups already selected to be restricted in general intelligence, self-report predictors, especially interest tests, are more valid predictors of graduate majors than are the cognitive tests. This is especially true for females, who were not so career oriented in 1960 as they are today. The difference is apparent in the ninth grade and continues to the twelfth as each set of tests becomes better at differentiating with respect to group membership. It is also clear that the cognitive set adds substantially to the predictive accuracy of self-report measures.

Not only did the restriction of range in general intelligence attenuate the predictive validity of the cognitive tests, but it also increased the validity of the

self-report measures. The cross-correlations between cognitive and self-report scores are so small in the full range of high school talent that dependence solely on interest tests for predicting group membership leads to many errors. The consideration of a cognitive floor for students being counseled on undergraduate and graduate majors is, therefore, essential.

Some Cautions

Choosing criterion groups. Throughout this chapter, we have argued for the utility of predicting group membership for assessing the differential validity of contrasting ability profiles. When predicting individual membership in groups, the definition of the group is critically important. Every member does not have to be successful and satisfied, but a given group should meet those criteria on average. If one has current information concerning proficiency in performance and satisfaction for the members of a group, this information can be used to select a more homogeneous subset of the group.

Not only should the members of the groups be reasonably successful and satisfied (Dawis & Lofquist, 1984; Lofquist & Dawis, 1991), but policymakers should be satisfied with the overall performance of current groups. The discriminant methodology involves placing the persons tested in advance of the formation of the group at an estimated difference from the mean of the group. It is important to keep in mind that persons can be either too high or too low on the combination of predictors to be well suited for membership in certain groups.

The predictive design. Experience in educational or occupational groups can change mean levels of scores on both cognitive and self-report tests. The high school curriculum chosen affects test scores. Thus, at least part of the gain in predictive accuracy from the ninth grade to the twelfth grade is due to differential curriculum choice. Out-of-school learning and psychological maturation surely are also involved. Thus, the predictive design is essential. A concurrent design may be less biased when the regression methodology is used, but concurrent validation does not possess the scientific significance of predictive validation over extended periods of time.

Not-So-Firm Inferences

Spatial ability and academics. The high school grade point average reported by students in the upper 20 percent in space and math was below what one would expect from their level of general intelligence, or even their level

of verbal ability. On the other hand, they reported approximately equivalent grades in math in comparison to the other two contrasted groups and were only a little below that level in science. They did, however, have the highest grades of the three groups in technical-vocational courses. Thus, generally low motivation to do well in school cannot explain these students' lower grade point average overall. What can explain it? We suggest that these students in the high space-math group may have been turned off by the verbal nature of the high school curriculum and the verbal nature of the tests used to assess achievement in school. No direct test of this hypothesis occurs to us, but some unpublished military data is suggestive. There is a modest positive correlation between scores on printed tests of proficiency in technical jobs and supervisory ratings of motivation and leadership. The latter are also correlated with the proficiency ratings of subordinates, which in turn are correlated with scores on proficiency tests. It is difficult to believe that a person can be a leader in a technical specialty without being proficient in that specialty. Thus, when test-methods scores are controlled in the proficiency tests, the correlations with the ratings increase. Similarly, when rating-methods scores are controlled in the motivation and leadership ratings, the correlations with the uncorrected proficiency tests also increase.

Changing competence in a group. The regression methodology readily allows the upgrading of competence in a group whenever there are sufficient applicants, but it also allows all too readily the selection of persons who are overqualified. This change can easily occur without planning. Upgrading the competence of a group by the discriminant methodology requires planning. It may be as simple as using current measures of proficiency and satisfaction to select the desired subsample. On the other hand, changing the nature of a group may require changing the precollege curriculum so to make it attractive to other individuals, as well as encouraging students to adopt different aspirations.

The data that we have presented may indicate the need for change in two of our groups. One group consists of the male doctorates in education; the other consists of the doctorates of both sexes in the social sciences. Both are concerned with the most important problems faced by our society. Males in the graduate education group anchored the low end of the scale on each of the first two cognitive functions. Females in education, on the other hand, had about average levels, among the graduate major groups, of abilities measured by the tests defining the functions. Yet the group of educators raised to positions of leadership has been predominantly male for many years.

Moreover, our data show that there is relatively little differentiation among

students choosing law, humanities, and the social sciences when using a wide selection of cognitive and self-report scores from high school. Some would argue that future social scientists should be less similar to the other members of that cluster and more similar to biologists, or even to physical scientists and engineers.

Use in selection. Use of the discriminant methodology for the selection of personnel in education, industry, and the military is clearly in conflict with the long history and well-established place of the regression methodology in personnel selection. A proposal to use the former methodology is novel in the experience of practically all personnel psychologists, but the proposal has some important advantages. Group membership is a cumulative criterion. Also, the combination of success and satisfaction in group membership merges two criteria that are often seen as relatively independent of each other (Dawis & Lofquist, 1984). The selection of competent persons who do not quickly create vacancies by their voluntary departure is eminently desirable.

Whether or not prediction of group membership is accepted, the usefulness in educational selection of a measure of spatial visualization is clearly indicated by our data. We do not expect all involved to jump at this suggestion on the basis of our data alone. Yet we feel that we have, at a minimum, provided the foundation for a large-scale field trial of an expanded college entrance test. The criteria in a field trial must involve more than academic grades if spatial ability is to emerge as an important predictor. This idea should not simply be dismissed as impracticable. The Educational Testing Service did introduce a third score in the Graduate Record Examination (GRE) just a few years ago that adds less, based on current evidence, to the regression validities of the verbal and quantitative scores than would a measure of spatial visualization. This measure also would add much less incremental validity, if any, to the prediction of group membership in educational and occupational groups compared to spatial visualization.

In 1999, an augmented version of the GRE is scheduled to appear. A two-track package will be offered, with four tests in both packages (three of which will be identical in each package). The identical tests will be verbal reasoning, analytical reasoning, and a writing exercise. For graduate programs in the humanities and the social sciences, the first package will include a quantitative test much like the current GRE-Q. In the second package, a mathematical-reasoning test predicated on knowledge through precalculus will be recommended to graduate programs in the physical sciences, mathematics, and engineering. This advanced quantitative test was designed to forestall GRE-Q

ceiling problems for the more technical disciplines. This might be an opportunity, however, for psychology to view itself as closer to the biological sciences than to the humanities, and, as such, require future graduate students to assimilate the conceptual tools of the natural science disciplines. Faculty members certainly will reveal their preferences for differential ability patterns by the tracks for which they lobby, but perhaps it would be most useful to require all psychology undergraduates to take all four tests—which is an option that would require a minimal investment of time.

It is interesting that the analytic test on the GRE is still retained, and spatial abilities are ignored; we are unaware of any evidence documenting the incremental validity of this measure over and above the information provided by the verbal and quantitative subtests for any discipline. For the departments of engineering and the physical sciences, a measure of spatial visualization would surely provide more useful information. Also, do such disciplines need two assessments of verbal ability, at the cost of not measuring spatial ability, or would one suffice?

Possible curricular changes. If some highly able students are turned off by the present high school curriculum and examining practices, as we suggest, changes are possible. For example, practically all laboratory scientists advocate more hands-on science experience from early grade school through high school. College preparatory courses in technology are possible. It is not a problem to pitch these at the level of ability required for success in college. In addition, foreign language courses that are less literary and more functional are possible. These can emphasize conversational competence, the reading of newspapers, and assignments in the foreign-language equivalent of *Scientific American*. English literature courses can be reoriented to include science fiction, biographies of scientists and mathematicians, and selected articles from *Scientific American*.

Some speculations. There are many adults who did poorly in school but were highly successful in science, technology, and business as well as many of the creative arts. Edison, Ford, and Langley may well have been spatially gifted individuals. There are probably many more such cases than there are cases of highly successful novelists, playwrights, essayists, and biographers who failed in school. It may not be necessary to invoke motivation, hard work, dependability, and other personality traits to explain such examples. Individual differences in spatial visualization may tell most of the story.

Our research suggests that the two intellectual cultures depicted by C. P.

Snow (1964) have a psychological reality. These two cultures are well established as early as the ninth grade, and the two developmental paths certainly mirror Vernon's *v:ed* and *k:m* constructs in many conspicuous ways. As children develop, differential experiences most likely augment this differentiation, but a person's pattern of abilities and preferences at age fourteen helps determine his or her subsequent experiences (Rowe, 1994; Scarr, 1992; Scarr & McCartney, 1983). It seems to us that our schools only do a good job of fostering development in one of these cultures—the more verbal one.

Conclusion

Students who are talented spatially are being overlooked by our educational system. One reason for this is the overreliance on letter grades in the highly verbal high school curriculum. A second, related reason is the overreliance on correlations between verbal and quantitative predictor tests, and the verbal and quantitative tests used to measure achievement in training and educational curricula. We suggest that educators rethink the dictum "If students can't write, they can't think." To be sure, this is true of some people, but they also tend to be low in spatial and mathematical-reasoning abilities.

Spatially talented persons prefer to solve problems involving ideation about things. They are turned off by abstract verbal subject matter. This does not mean, however, that their becoming mechanics, skilled workers, or technicians is a satisfactory solution to the career dilemma for people at all ability levels. Those students in the upper 20 percent of our space-math group were still well above average in verbal ability and should have been encouraged more strongly to prepare for professional careers. But these students are currently being excluded from the most select institutions for advanced training in engineering and the physical sciences.

Curriculum-adjustment interventions are possible ways of salvaging intellectually talented students who are not so verbally able as they are in other intellectual arenas. Precollege science should contain substantial hands-on experience. It would also be useful if engineering professionals would undertake to prepare a technological sequence for the college-preparatory curriculum in high school. Less than 200 years ago, the physical sciences entered high school and college curricula. Less than 150 years ago, agriculture, military science, and the mechanical arts entered the postsecondary curriculum through the opening created by an act of Congress. It took many more years for the land-grant colleges to become respectable in the eyes of most academics. It is past time for more change.

References

- Benbow, C. P. (1988). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. *Behavioral and Brain Sciences* 11:169–232.
- Benbow, C. P. (1992). Academic achievement in mathematics and science of students between ages thirteen and twenty-three: Are there differences among students in the top one percent of mathematical ability? *Journal of Educational Psychology* 84:51–61.
- Benbow, C. P., & Lubinski, D. (1994). Individual differences amongst the mathematically gifted: Their educational and vocational implications. In N. Colangelo, S. G. Assouline, & D. L. Ambrosio (Eds.), *Talent development* (Vol. 2, pp. 83–100). Dayton: Ohio Psychology Press.
- Benbow, C. P., & Stanley, J. C. (1982). Intellectually talented boys and girls: Educational profiles. *Gifted Child Quarterly* 26:82–88.
- Benbow, C. P., & Stanley, J. C. (Eds.). (1983). *Academic precocity: Aspects of its development*. Baltimore: Johns Hopkins University Press.
- Binet, A., & Simon, T. (1905). *Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux*. *L'année psychologique* 11:191–244.
- Burt, C. (1940). *The factors of the mind*. London: University of London Press.
- Cleary, T. A., Humphreys, L. G., Kendrick, S. A., & Wesman, A. (1975). Educational uses of tests with disadvantaged students. *American Psychologist* 30:15–41.
- Dawis, R. V. (1991). Vocational interests, values, and preferences. In M. D. Dunnette & L. M. Hough (Eds.), *The handbook of industrial/organizational psychology* (2nd ed., Vol. 2, pp. 833–871). Palo Alto: Consulting Psychologists Press.
- Dawis, R. V., & Lofquist, L. H. (1984). *A psychological theory of work adjustment*. Minneapolis: University of Minnesota Press.
- Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist* 43:95–103.
- 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.
- 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. (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. (1991a). Limited vision in the social sciences. *American Journal of Psychology* 104:333–353.
- Humphreys, L. G. (1991b). 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. (1992). Commentary: What both critics and users of ability tests need to know. *Psychological Science* 3:271–274.
- Humphreys, L. G., Davey, T. C., & Kashima, E. (1986). Experimental measures of cognitive privilege/deprivation and some of their correlates. *Intelligence* 10:355–370.
- Humphreys, L. G., Lubinski, D., & Yao, G. (1993). Utility of predicting group membership and the role of spatial visualization in becoming an engineer, physical scientist, or artist. *Journal of Applied Psychology* 78:250–261.
- Humphreys, L. G., & Yao, G. (Unpublished manuscript.) *Predicting graduate majors*. University of Illinois at Champaign.
- Hunter, J. E. (1983). *Test validation for twelve thousand jobs: An application of job classification and validity generalization analysis to the General Aptitude Test Battery* (Test Res. Rep. No. 45). Washington, D.C.: Department of Labor, U.S. Employment Services.
- Lofquist, L. H., & Dawis, R. V. (1991). *Essentials of person-environment-correspondence counseling*. 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. (1994). The Study of Mathematically Precocious Youth: The first three decades of a planned fifty-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, N.J.: Ablex.
- Lubinski, D., Benbow, C. P., & Sanders, C. E. (1993). Reconceptualizing gender differences in achievement among the gifted. In K. A. Heller, F. J. Mönks, & A. H. Passow (Eds.), *International handbook of research and development of giftedness and talent* (pp. 693–707). Oxford: 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: 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.
- McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. A. (1990). Project A validation results: The relationship between predictor and criterion domains. *Personnel Psychology* 43:335–353.
- McNemar, Q. (1942). *The revision of the Stanford-Binet scale*. Boston: Houghton Mifflin.
- Paterson, D. G. (1938). The genesis of modern guidance. *Educational Record* 19:36–46.
- Paterson, D. G., Elliott, R. M., Anderson, L. D., & Toops, H. A. (1930). *Minnesota mechanical ability tests*. Minneapolis: University of Minnesota Press.
- Prediger, D. J. (1976). A world-of-work map for career exploration. *Vocational Guidance Quarterly* 24:198–208.
- Rowe, D. C. (1994). *The limits of family influence*. New York: Guilford.
- Scarr, S. (1992). Developmental theories for the 1990s: Development and individual differences. *Child Development* 63:1–19.
- Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype environment effects. *Child Development* 54:424–435.
- Schmid, J., & Leiman, J. (1957). The development of hierarchical factor solutions. *Psychometrika* 22:53–61.
- Smith, I. M. (1962). *Spatial ability: Its educational and vocational significance*. San Diego: Knapp.
- Spearman, C. (1904). “General intelligence,” objectively determined and measured. *American Journal of Psychology* 15:201–292.
- Spearman, C. (1914). Theory of two factors. *Psychological Review* 21:101–115.
- Snow, C. P. (1964). *The two cultures and a second look*. Cambridge: Cambridge University Press.
- Stanley, J. C., Benbow, C. P., Brody, L. E., Dauber, S., & Lupkowski, A. E. (1992). Gender differences on eighty-six nationally standardized aptitude and achievement tests. In N. Colangelo, S. G. Assouline, & D. L. Ambrosio (Eds.), *Talent development: Proceedings from the 1991 Henry B. and Jocelyn Wallace National Research Symposium on Talent Development* (pp. 42–65). Unionville, N.Y.: Trillium Press.
- Terman, L. M. (1916). *The measurement of intelligence*. Boston: Houghton Mifflin.
- Terman, L. M., & Merrill, M. A. (1937). *Measuring intelligence*. Boston: Houghton Mifflin.
- Thorndike, R. L. (1994). *g*. *Intelligence* 19:145–155.
- Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs* (No. 1). Chicago: University of Chicago Press.
- Vernon, P. E. (1947). Research on personnel selection in the Royal Navy and British Army. *American Psychologist* 2:35–51.
- Vernon, P. E. (1950). *The structure of human abilities*. New York: Wiley.
- Wechsler, D. (1941). *The measurement of intelligence*. Baltimore: Williams and Wilkins.
- Wise, L. L., McLaughlin, D. H., & Steel, L. (1979). *The Project TALENT data bank*. Palo Alto: American Institutes for Research.

III

What Do We Know about Proper Provisions for the Gifted?

In the previous section, we were left with a disappointing message: social science has produced a great deal of knowledge during its first century, but not much of this knowledge is used well. Therefore, the role that social science does and can play in the betterment of society is compromised. Indeed, if we were to decide to apply appropriately the knowledge that social science has generated, we could see a significant enhancement of societal functioning. This leads to some compelling questions: What do we know? What does work? What could we do? In this section we attempt to answer these questions as they pertain to the education of gifted children—children whose intellectual competence is so advanced for their age that the regular curriculum does not and cannot meet their educational needs.

These questions are perhaps especially critical now, because the current school-reform effort and economic considerations threaten the viability of the very programs designed to serve gifted students (Benbow & Stanley, in press). Across the country, programs for the gifted are being closed down or scaled back dramatically. This is truly a time of shrinking resources; that makes it especially important to know how to use in the most effective manner whatever resources are available. Sadly, in many programs for the gifted we have not been doing so; we have not been using our knowledge well. Such programs often have consisted of “fun and games” that could be, and often were, argued to be beneficial to all students. The programs were not designed to meet the specific educational needs of gifted students; therefore, they are difficult to justify and to defend against cuts. Consequently, school reform, which is of great concern to the gifted-education community, could be beneficial. It might force us to pause and to reflect on and evaluate our practices; it could be the stimulus for