Incorporating General Intelligence Into Epidemiology and the Social Sciences

DAVID LUBINSKI

Iowa State University

LLOYD G. HUMPHREYS

University of Illinois

The purpose of this article is to demonstrate the value of examining a variety of pressing behavioral, medical, and social phenomena as they relate to gradations in general intelligence. Although few (if any) variables in the social sciences can compete with the construct of general intelligence in its ability to forecast an array of socially valued attributes and outcomes, measures of general intelligence are seldom incorporated into correlational and experimental designs aimed at understanding *maladaptive behavior* (e.g., crime, dropping out of high school, unwise financial planning, health-risk behaviors, poor parenting, and vocational discord) or its opposite, highly *adaptive behavior.* We contend that, if consulted more often, the construct of general intelligence would contribute to understanding many puzzling human phenomena, because successive gradations of intelligence reflect successive degrees of risk. A method is provided for uncovering group trends, one expressly designed to reveal the range and prevalence of the many different kinds of human phenomena that vary as a function of intellectual gradations. By employing this method, policymakers and the public can more readily apprehend the significant, but often unsuspected, contribution made by general intelligence to many socially important outcomes. Our approach is similar to traditional epidemiological research aimed at ascertaining antecedents to maladies through the defining features of high-risk groups (e.g., for lung cancer, smokers and passive smokers; for AIDS victims, participants in unsafe sex; for academic mediocrity, among the intellectually gifted in nonaccelerative educational tracks; for mental retardation, high blood-lead levels). Once such high-risk groups are defined (i.e., groups of persons whose behavioral dispositions predispose them, and often others around them, to unfortunate outcomes), policymakers and scientists are in a better position to disentangle genuine causes from families of correlations and can concentrate ameliorative resources more effectively. Data from educational and medical contexts are analyzed to show how measures of general intelligence, and other dimensions from differential psychology, can complement epidemiological and social science inquiry. We also argue that by incorporating such measures of human variation into policy development and research, policymakers are more likely to forestall "iatrogenic effects" (maladies caused by treatment).

The scientific significance of general intelligence is underappreciated in epidemiology and the social sciences. Its potential for informing public policy is grossly

underappreciated. In part, this neglect is because intellectual assessment has typ-ically focused on individuals (a hallmark of the psychometric approach) as dis-tinct from populations (more of an epidemiological approach). One typically examines group trends when forming public policy (e.g., when launching public health campaigns). This article illustrates how taking more of an epidemiological approach to the role of intelligence (and other psychological traits) in social mal-adies promises to clarify the development and treatment of many pressing social problems. Teaming measures of general intelligence with existing concepts and methods in epidemiology and the social sciences will not only sharpen scientific inquiry, it also will serve to enhance public policy development and research.

Our treatment does not attempt to systematically review current epidemiologi-cal applications or the array of analytic tools at present available in differential (individual differences) psychology. Rather, we seek to make a more general point by combining key aspects of each: Our aim is to communicate why both epidem-iology and the social sciences are likely to profit from assessing general intel-ligence, incorporating such measures into preexisting research designs, and tailoring preventive measures and interventions toward individual differences un-covered by these assessments. If this article motivates readers to consider the role general intelligence might play in determining epidemiological and social science outcomes relevant to public policy formation and implementation, we will have achieved our intended goal.

GENERAL OVERVIEW

Unlike psychometrics, epidemiological research is not concerned primarily with individual differences in treatments and outcomes. As the word implies, from its Greek etymology, epidemiology means knowledge "upon the population." More-over, epidemiology is concerned not only with the causes of pathology but also with identifying populations in which those causes are most often likely to be found. Incidence and prevalence rates of maladies within various high-risk groups often uncover risk factors conducive to their contraction and spread. Those factors sometimes consist of behaviors, as in the AIDS epidemic. It was not that high-risk groups possessed elevated host susceptibilities for HIV infection, it was rather the *behaviors* of the high-risk groups that placed them in harm's way. By better defin-ing what it is about these groups that places them at risk (through behaviors, as opposed to gross demographic categories), intervention strategies appropriately shifted from risks associated with group membership (e.g., gay males, IV drug users) to risks associated with high-risk behaviors (unsafe sex, sharing needles). This discovery enabled health care professionals to launch more effective inter-ventions aiming preventive measures at specific behaviors rather than at specific groups.

In epidemiology and the social sciences, a variety of behavioral attributes have been used to further specify particular aspects of groups that constitute risk. In addition to degrees of participation in unsafe sex and sharing needles, groups have been formed based on level of other habits (smoking, diet, exercise) and broad personality dimensions routinely assessed in differential psychology. Here, we focus on the construct general intelligence because, for behaviors and outcomes that interest most psychologists and policymakers, it manifests the most impressive array of external correlates relative to other psychological dimensions—and relative to other ways of aggregating people. It is not everything; nothing is. Almost all complex forms of human behavior have multiple determinants. Nevertheless, general intelligence has been grossly underappreciated in social science research (Humphreys, 1991; Meehl, 1990).

Although we do not embrace all of the arguments found in Herrnstein and Murray's book, *The Bell Curve* (1994), by illustrating how general intelligence fares when it competes with conventional explanatory social science variables, such as SES, in forecasting many different kinds of important social behaviors and outcomes (such as academic achievement, crime, poverty, poor parenting), these authors have made an important contribution (cf. Bouchard, 1995). They also highlighted the need to examine SES and general intelligence conjointly; and, through systematic documentation, they convincingly showed "in table after table, graph after graph, that cognitive ability has become a more important determinant of social status than social class of origin" (Bouchard, 1995, p. 416).

Differential psychologists were not surprised by the empirical findings reported in Herrnstein and Murray (1994). Most have been reported in traditional outlets for decades (Anastasi, 1958; Brody, 1992; Cronbach & Snow, 1977; Eysenck, 1979, 1995; Gottfredson, 1986; Gottfredson & Sharf, 1988; Jenkins & Paterson, 1961; Jensen, 1980; Loehlin, Lindzey, & Spuhler, 1975; Tyler, 1965; Vernon, 1961, 1979; Willerman, 1979). But many social scientists and policymakers were surprised, as was the public. Now that the scientific significance of general intelligence is more fully appreciated (Bouchard, 1995; Gottfredson, 1997a, 1997b; Snyderman & Rothman, 1987, 1988), it behooves us to consult this dimension more often and with more scientific systematicity. We need to assess general intelligence for its possible incremental validity, along with whatever else we assess, for better understanding the pressing social problems currently facing us.

Our reading of mainstream science on intelligence has revealed a convergent and discriminant patterning consistent with the following working hypotheses: general intelligence functions as a proximate causal determinant for an individual's exposure, host suitability, and likelihood to develop an array of socially desirable and undesirable behaviors and outcomes (including crime, educational

credentials, HIV infection, income, and parenting quality). (We recognize that intelligence develops on polyenvironmental and polygenetic substrates that represent more distal causes.)

Childhood assessments of general intelligence are consistently predictive of a broad spectrum of high-risk maladaptive as well as adaptive behaviors (Brand, 1987; Eysenck, 1979, 1995; Herrnstein & Murray, 1994; Jensen, 1980): delinquent behavior, rate of learning, high-risk health behaviors, knowledge and skills pertinent to caring for the physical and psychological well-being of children. These covariates translate into a family of socially undesirable risks for populations within lower segments of the intelligence dimension; or, conversely, a family of socially desirable assets for populations within higher segments (Herrnstein & Murray, 1994). These behaviors and outcomes associated with differential gradations of general intelligence are important, not only for understanding these phenomena from a scientific point of view, but also for developing public policy and creating effective interventions. Other dimensions of individual differences will likely also play a role in understanding important social phenomena; however, based on our reading of the empirical evidence, our hypothesis is that the greatest single contribution will be made by general intelligence.

In the following sections, we first briefly review the ways in which general intelligence may be assessed (because it can be evaluated in more ways than many people realize). Second, we offer a method for uncovering group trends as a function of continuous intellectual gradations (consistent with more of an epidemiological approach designed to inform public policy about likelihoods of certain outcomes for populations at risk). Third, we present and analyze data on some important social phenomena and make suggestions for badly needed research. We also suggest how this approach might contribute to conventional, ongoing epidemiological research in medical science. Finally, our discussion will conclude by introducing Carnap's (1950) "total evidence rule" and its bearing on a fallacy in logic, the fallacy of the "neglected aspect" (Castell, 1935). Both concepts illustrate that, indeed, it is hard to justify excluding intelligence from research on health and behavior.

MANY WAYS OF ASSESSING GENERAL INTELLIGENCE

The sheer number of distinct kinds of intelligence tests available today (Carroll, 1993) gives rise to a number of concerns about the assessment of general intelligence. Many of these concerns reflect a dated form of "literal operationism," to wit, each distinct measure necessarily taps somewhat distinct intellectual functions. But this expectation rarely holds in science. It is not true for measures purporting to assess atmospheric temperature with different materials, for example, and it is certainly not true for psychological measures of behavioral attributes. Many psychological measures with different names and distinct items (such

as academic ability, aptitude, scholastic ability, scholastic achievement) can, and often do, measure essentially the same thing (Cleary, Humphreys, Kendrick, & Wesman, 1975; Lubinski & Dawis, 1992; McNemar, 1964). From a scientific point of view, a measure's meaning (technically, its construct validity) is found in its network of causes and correlates, not in the unique aspects of its item content or label. Precisely because general intelligence is general, it can be measured in any of several ways, just as temperature can be measured with a number of different chemical materials. Finding material differences in assessment tools does not imply that they assess different dimensions of human variation. An example followed by an empirical demonstration may be useful here.

If the items of a test covary with each other, say an average intercorrelation of .25, and the remainder of their composition is nonoverlapping with other items in the array (each holds about 94% uniqueness, independent of the others), aggregating the set of item scores forms a composite measure for indexing individual differences that is primarily defined by the bits of communality running across each (cf. Gordon, 1997). This is readily shown by one of the most celebrated formulas in psychometrics, the Spearman-Brown Prophecy formula: common variance $= k(r_{xx}) \div [1 + r_{xx}(k - 1)]$, where $k =$ number of items, and $r_{xx} =$ the average intercorrelation of each item. For example, aggregating 30 vocabulary items whose average interitem correlation is .25 results in a composite score approximately 91% determined by the communality running through the items (the remaining 9% being determined by random measurement error). The presence of substantial uniqueness in each and every item cannot be avoided, but when even a small amount of communality exists within each item, aggregating a sufficient number attenuates the uniqueness of each to a tiny sliver of systematic bias. The net result is an internally consistent measure of an important dimension of human functioning subject to a small amount of measurement error (Green, 1978).

Further, ostensibly quite different tests may generate interchangeable patterns of external correlations, which would justify their use as measures of the same construct, equivalent for purposes relating to those correlations. When this happens, the veins of research generated by each measure may be combined to form a cohesive body of knowledge about the construct under analysis. Consider the following.

We have elsewhere stressed that the construct of general intelligence is best assessed by including content cutting across verbal, quantitative, and pictorial domains (Humphreys, 1979, 1994; Humphreys, Lubinski, & Yao, 1993a; Lubinski & Dawis, 1992). Each of these domains is multifaceted and may be splintered into more molecular strands. Our illustration begins by focusing on verbal ability, which is highly correlated with general intelligence. Measures of verbal ability have ranged from "reading comprehension" to "literary information" to "vocabulary" (e.g., antonyms, synonyms, analogies), and there are others. Table 1 is from Lubinski and Dawis (1992, p. 22). It examines the patterns of external

164   LUBINSKI AND HUMPHREYS

TABLE 1
Profiles of Correlations of Various Other Measures Across Three
Measures of Verbal Ability

	Literature	Vocabulary	Reading Comprehension
Aptitude Tests			
Mechanical reasoning	.43	.52	.54
2-D visualization	.25	.32	.35
3-D visualization	.35	.43	.47
Abstract reasoning	.45	.53	.61
Arithmetic reasoning	.54	.63	.63
High school math	.57	.59	.57
Advanced math	.42	.43	.39
Information Tests			
Music	.67	.68	.62
Social studies	.74	.74	.71
Mathematics	.62	.63	.57
Physical science	.64	.67	.60
Biological science	.57	.61	.56
Interests			
Physical sciences	.24	.25	.22
Biological sciences	.26	.25	.22
Public service	.16	.12	.12
Literary–linguistic	.37	.32	.32
Social service	.07	.06	.07
Art	.32	.30	.29
Music	.23	.20	.20
Sports	.12	.12	.13
Office work	−.35	−.29	−.27
Labor	−.08	−.06	−.06

Note. These correlations were based on female subjects (male profiles are similar). $N = 39,695$. Intercorrelations for the three measures were the following: literature/vocabulary = .74, literature/reading comprehension = .71, and vocabulary/reading comprehension = .77. Reprinted with minor alterations from Lubinski & Dawis (1992, p. 22) by permission of the publisher.

correlates of three such measures over the full range of talent ($N = 39,696$ 10th grade students, taken from a stratified random sample of U.S. high schools compiled by Project Talent, Flanagan et al., 1962): literature information (24 items), vocabulary (30 items), and reading comprehension (48 items). Across an array of aptitude and information tests as well as educational/vocational interest questionnaires, the patterns of external correlations for all three verbal ability measures are essentially functionally equivalent; that is, empirical evidence indicates inferences based on one would be highly compatible with those derived from the other two. By correlating each column of correlations in Table 1 with the other two, the three resulting correlations range from .97 to .99. This is what is meant by functional equivalency of psychological measures: two measures (which may look

quite different and carry quite different names) assess individual differences that translate into interchangeable forecasts over a broad mix of criteria. As in this example, although these three verbal ability measures have different names and distinct content, individual differences assessed by each nevertheless translate into *congeneric inferences* over many and diverse criteria.

Notice also the high correlations between these verbal ability measures and quantitative and spatial/mechanical reasoning abilities and information tests. A similar positive manifold (positive intercorrelations across measures) cuts across all heterogeneous collections of cognitive tests, provided a wide range of talent is sampled (as done here). The communality running through such positive manifolds constitutes the psychometric basis for defining general intelligence. This general dimension reflects overall complexity or sophistication of the intellectual repertoire, and the profile of major markers of subsidiary dimensions (mechanical, quantitative, spatial, and verbal content) reflects more specific strengths and weaknesses of individuals (Humphreys, 1962, 1979; Lubinski & Dawis, 1992), which are frequently important to measure as well (Humphreys et al., 1993a).

Just as different aggregations of unique kinds of verbal materials may generate interchangeable correlational profiles, different aggregations of verbal, quantitative, and spatial-mechanical content (different tests of general intelligence) also generate interchangeable profiles. One can even assess the construct of general intelligence using variegated conglomerations of general information tests (e.g., composites consisting of information about sports, home economics, animals, architecture, the military, the Bible, journalism, sedentary games). But the individual items on such tests are so lightly correlated with the general factor of intelligence that it requires a larger number of questions to build a scale whose pattern of external correlates approximately parallels those of traditional measures of general intelligence (Roznowski, 1987). In contrast, typical items on general intelligence tests are more focused on verbal/linguistic, spatial/pictorial, quantitative/mathematical reasoning and, as such, are more closely associated with cognitive complexity than are questions about information. Nevertheless, such demonstrations reinforce the fact that we are talking about a robust phenomenon, capable of being measured in several ways (Jensen & Weng, 1994). They are thus also consistent with Spearman's (1927) phrase "indifference of the indicator," illustrating the very general psychometric phenomenon he had in mind. Namely, the measurement of general intelligence does not depend on items limited to any specific kind of content (or presentation modality: oral/written, individually/group administered).

This discussion, we hope, will forestall concerns about different types of content found on contrasting general intelligence tests (or other psychological measures purporting to assess the same construct in different ways), which are a frequent source of confusion concerning the psychometric approach. Differential test content does not necessarily equate with distinct dimensions of individual

differences that translate into dissimilar forecasts. Whether they do or not is an empirical question. The correlational patterns of well-developed general intelligence tests (like the aforementioned verbal ability measures) are typically quite similar. There are many more or less equivalent vehicles for studying general intelligence. And many large social science data banks contain good measures (or scales which may be aggregated to form acceptable measures) of general intelligence.[1]

INCORPORATING INTELLIGENCE AND OTHER INDIVIDUAL DIFFERENCES INTO POLICY-RELATED RESEARCH: PRESENTING DATA AND INTERPRETING FINDINGS

Because the behaviors and outcomes that interest policymakers are profoundly multiply determined, multiple predictors are often useful to assemble for policy development and research. Typically, any given variable tells only part of the story. Moreover, the same is true for evaluating policy outcomes, because the implementation of a particular policy frequently engenders multiple effects not limited to the more circumscribed goals of immediate concern to policymakers. So policy research is inherently multivariate on both (input and output) sides of the equation. And it is often useful to evaluate the relative contribution of different predictors, especially when planning interventions. In addition, one must be vigilant not to dismiss ostensibly small functional relationships, because they can have appreciable utilities. Communicating their substantive significance, however, requires a method for doing so. The following pages discuss these and other

---

[1]For a number of purposes, narrow estimates of intelligence such as verbal comprehension or nonverbal reasoning (Raven's matrices) can be used interchangeably, but for some research purposes a truly general test is required. For example, a good deal of skepticism about intelligence tests has been expressed because of reported gains in means in raw scores extending over several decades, the so-called Flynn effect (Flynn, 1984, 1987). Gains in Raven's matrices are larger than those reported on verbal tests, which are larger than gains on measures that sample more widely (across quantitative, spatial–mechanical, and verbal reasoning content). An unexplored hypothesis emerges from these findings. Narrow estimates of the general factor have about 50% of their raw score variance accounted for by the general factor, another 40% by nonerror uniqueness, leaving 10% random error. In contrast, a broad test of general intelligence can have 85% of its raw score variance accounted for by the general factor, 5% nonerror uniqueness, and 10% random error. Furthermore, the nonerror uniqueness is most homogeneous in content and operations for the Raven, somewhat more heterogeneous for verbal tests, and highly heterogeneous for a broad test of general intelligence. The question is, What proportion of the reported gains are associated with the general factor versus the nonerror uniqueness? We do not doubt that gains can take place on the general factor as a result of cultural change over decades, but it is quite possible that the reported gains are the result of increases in scores on nonerror unique components of variance as well as construct-valid components of the general factor. The ordering of the size of the Flynn effect does seem to correspond inversely with the size of the general factor variance. Nevertheless, given that structural relationships between individual-differences variables are more resistant to change than are mean levels, changes in means are not likely to influence either the slope of a regression line or construct validity.

key issues in incorporating intelligence and other individual differences into poli-
cy-related research.

Value of Using Multiple Predictors and Representative Samples

Although adding measures of general intelligence to epidemiological and social
science research would enhance our understanding of many different kinds of
human phenomena, an ideal approach also would include other dimensions from
differential psychology: personality (Eysenck, 1995; Lykken, 1995), vocational
preferences (Dawis, 1991, 1992), and other abilities (Humphreys et al., 1993a;
Lubinski & Dawis, 1992). A good approximation of the kind of study, which one
might aim for, is Project Talent (Flanagan et al., 1962),[2] a longitudinal study that
was begun in the late 1950s. Project Talent consists of a stratified random sample
of the nation's high schools, grades 9 through 12. Students were assessed on a
wide range of aptitude, achievement, and information tests, in addition to educa-
tional and vocational interests and personality questionnaires. (All the data report-
ed in Table 1 are from Project Talent.) Students also completed a comprehensive
394-item background questionnaire, which included questions about family,
home, health, hobbies, personal experiences, and future plans. Follow-up ques-
tionnaires were secured from students at one, five, and 11 years following their
high school graduation. Readers interested in the particulars of an excellent data-
bank model, which might be teamed with concepts and methods from epidemiol-
ogy and the social sciences, would do well to consult Flanagan et al. (1962) and
Wise, McLaughlin, and Steel (1979).

There was a problem securing representative samples for Project Talent's three
follow-up surveys, however; but sampling weights based on the initial sample of
participants and for those traced longitudinally are available (Wise et al., 1979).
Nevertheless, with modern augmentations (including more systematic longitudi-
nal tracking), and a few additional outcome criteria, Project Talent offers an excel-
lent model for gaining a purchase on other individual differences variables likely
to complement existing measures in epidemiology and the social sciences (some
examples are provided here). It provides a huge stratified random sample, so
restriction of range on purportedly key determinants, which serves to attenuate the
magnitude of observed functional relationships (Brand, 1987), is not an issue.
Moreover, participants in Project Talent were longitudinally tracked (over three

---

[2]Because of its comprehensiveness and depth, we have found this data bank to be unique for
enabling investigators to address sources of environmental privilege–deprivation (Humphreys,
Davey, & Kashima, 1986), developmental disabilities (Humphreys, Lubinski, & Yao, 1993b), intel-
lectual versus SES covariates of health and physical well-being (Lubinski & Humphreys, 1992),
assessing cognitive abilities beyond general intelligence (Humphreys, Lubinski, & Yao, 1993a),
mathematical giftedness (Lubinski & Humphreys, 1990a), and testing theoretical hypotheses about
trait-by-trait interactions (Lubinski & Humphreys, 1990b). We have published Project Talent's entire
43-item health questionnaire in *Intelligence,* see Appendix in Lubinski and Humphreys (1992,
pp. 113–115).

time frames) and initially assessed on many different kinds of categorical and continuous predictors and (outcome) criteria—the breadth of which provides many opportunities for examining multiple outcomes simultaneously.

Value of Using Multiple Outcomes

When evaluating policy or the likely results of a proposed policy, it is useful to assimilate multiple outcome measures. Doing so reveals the generality of E.L. Thorndike's (1911) observation that "good things tend to go together," as do bad things. People at risk for a particular physical or social malady frequently are at risk for others. Participants in unsafe sex, for example, are not only at risk for AIDS but also a variety of other sexually transmitted diseases, just as people who share unsterilized needles are at risk for hepatitis as well as multiple infections of the blood—in addition to AIDS. Similarly, smokers place themselves at risk not only for major medical conditions such as lung cancer and emphysema, but also chronic bronchitis, susceptibility to colds and upper respiratory infections, work absenteeism, and (for pregnant women who smoke) a greater likelihood of delivering low-birth-weight babies. All three of these specific behaviors—unsafe sex, sharing unsterilized needles, and smoking—place individuals at risk for an array of undesirable conditions. Thus, when examining policy-relevant scientific evidence, it is ideal to have data from a wide network of relevant criteria to fully appreciate the multiple risks associated with a particular characteristic, behavior, or social policy.

Clarifying the Impact of Individual Differences: Graphing Group Trends and Correlating Group Means

For policy research and development, data need to be organized and presented in the clearest light possible. In epidemiological research, this typically involves analyzing data on groups, as opposed to computing correlations on individuals, because epidemiological approaches focus on inferences "upon the population." In this way, clear-cut trends within populations are more readily seen and communicated to policymakers and the public. For example, a recent special issue of the American Psychological Society's (APS) *Observer* (April 1996), on the Human Capital Initiative, contains a number of figures and tables of health criteria that vary as a function of traditional epidemiological and social science categories and continua (education, income, SES, gender, smoking behavior). Figure 1 is reproduced from this issue of the *Observer* (p. 22). It illustrates how the broad criterion "fair or poor health" varies as a function of seven levels of education. This graph also illustrates that the trend is not linear; risk increases dramatically with less than 12 years of education, with risk ratios for those with less than 12 years being from 1.5 to 2.5 computed relative to those with at least some college.

Herrnstein and Murray (1994) aggregated much of their data by successive gradations of general intelligence and SES, yet they have been criticized for cate-

Prevalence of fair or poor health as a function of education level

(National Health Interview Survey; adapted from Series 10, No. 179)
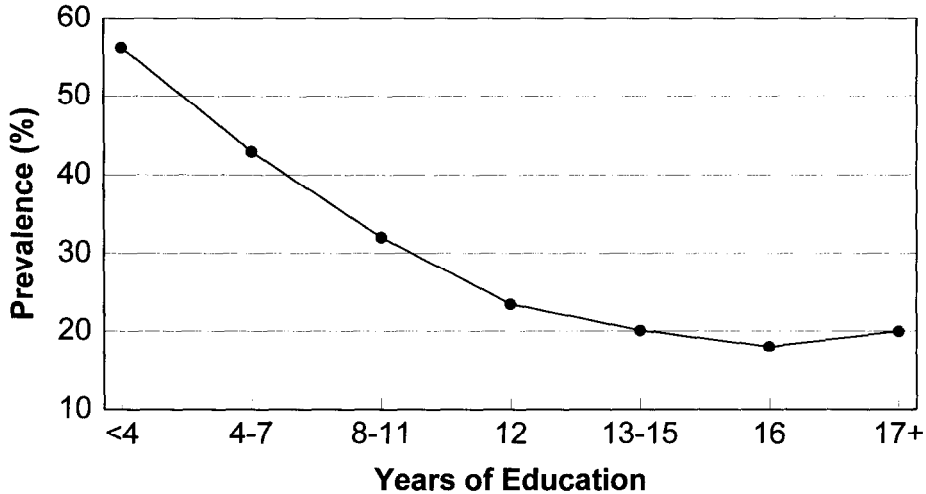

Figure 1. Reproduced from April 1996 American Psychological Society's *Observer* (p. 22).

gorizing continuous data, even though doing so is a standard procedure in epidemiology (Friedman, 1994) and the social sciences (Adler et al., 1994). They also have been criticized because the outcomes they plotted as a function of general intelligence and SES displayed only light to moderate correlations with these two variables, typically between .20 and .40 (e.g., Gould, 1994). Yet, individual-level correlations of this magnitude have profound social policy implications (cf. Rosenthal, 1990), as correlations between smoking and various health-related criteria routinely evince (Friedman, 1994).

Epidemiological studies have shown that smoking puts a developing fetus at risk for low birth weight (National Center for Health Statistics, 1994a)—a conclusion that few medical scientists would argue with today. Yet, if one computes a correlation (specifically, a phi coefficient) on the available data to quantify the magnitude of this relationship (Lubinski & Humphreys, 1996), the correlation is around .10. A phi coefficient was computed because the epidemiological studies in question collected data by dichotomizing birth weight (low versus normal) *and* smoking behavior (smoker vs. nonsmoker), when, in reality, both are continuously distributed variables with the latter being highly skewed. If continuous data were available, the usual Pearson correlation between smoking behavior and birth weight would likely approximate .25, a value approximating personal risk for a variety of pathologies as a function of smoking (we suspect risks associated with other habits to be of similar magnitude: e.g., alcohol consumption, sun tanning, consuming steroids, not wearing seat belts). This ambitious health campaign to

get people to quit smoking is thus based on correlations in the .20s. Unfortunately, many psychologists reflexively dismiss such values as trivial (epidemiologists would not).

These important risks are difficult to see from a scatter plot or even a correlation for individuals drawn from a stratified random sample across the United States, because birth weight and general health variations within continuous gradations of smoking behavior are so large (Lubinski & Humphreys, 1996). What is needed to highlight such important trends is a different approach. We suggest a graphic approach, an approach familiar to epidemiological and public health researchers—and similar to Figure 1. Our recommendation also will be familiar to experimental psychologists, who routinely plot, on some dependent variable, means that vary as a function of two or more treatments. (Individual differences measures also may be used to generate ROC curves familiar to signal detection theorists, for quantifying the sensitivity and specificity of individual differences variables, Humphreys and Swets, 1991).

Figure 2 shows how this application may be extended to individual differences variables using general intelligence. Figure 2 graphs two functional relationships between a garden-variety measure of general intelligence and two of its covariates, SES of origin and a general information composite.[3] But instead of simply providing two scatter plots or simple regression lines based on each individual participant, we have done something different. Just as epidemiologists and social scientists routinely divide SES into gradations, we have divided the general intelligence measure into 20 "intellectual gradations" (each defined by successive increments of .20 SD, X, extending from the mean of general intelligence in both directions). Then, within each X interval, we computed bivariate means to reveal the magnitude of change in the criterion variable across successive levels of intelligence. Correlations were computed on both sets of bivariate means, and both

---

[3]A detailed description of the composition and range of these measures is necessary to fully appreciate the story told by the covariation between intelligence and SES and intelligence and the general information composite: TALENT's intelligence composite is composed of tests of reading comprehension, arithmetic reasoning, and abstract reasoning (yielding a possible point range of 283). Project TALENT's intelligence composite comes very close to matching the content found on traditional measures of general intelligence, such as the Stanford-Binet scale (Terman & Merrill, 1960) and the Wechsler (1974) tests of general intelligence. TALENT's SES composite (135 possible point range) covers the conventional categories of economic/home/educational deprivation and privilege, namely, family income, value of home, number of books in the home, number of home appliances, quality of home's area for study, father's occupation and education, and mother's education; and finally, the general information composite (148 items), including art (12 items), law (9 items), medicine and health (9 items), engineering (6 items), architecture (6 items), journalism (3 items), foreign travel (5 items), military (7 items), accounting and business, sales (10 items), practical knowledge (4 items), clerical (3 items), Bible (15 items), colors (3 items), etiquette (2 items), hunting (5 items), fishing (5 items), outdoor activities (9 items), photography (3 items), sedentary games (5 items), theater and ballet (8 items), foods (4 items), miscellaneous (10 items), and general vocabulary (9 items).
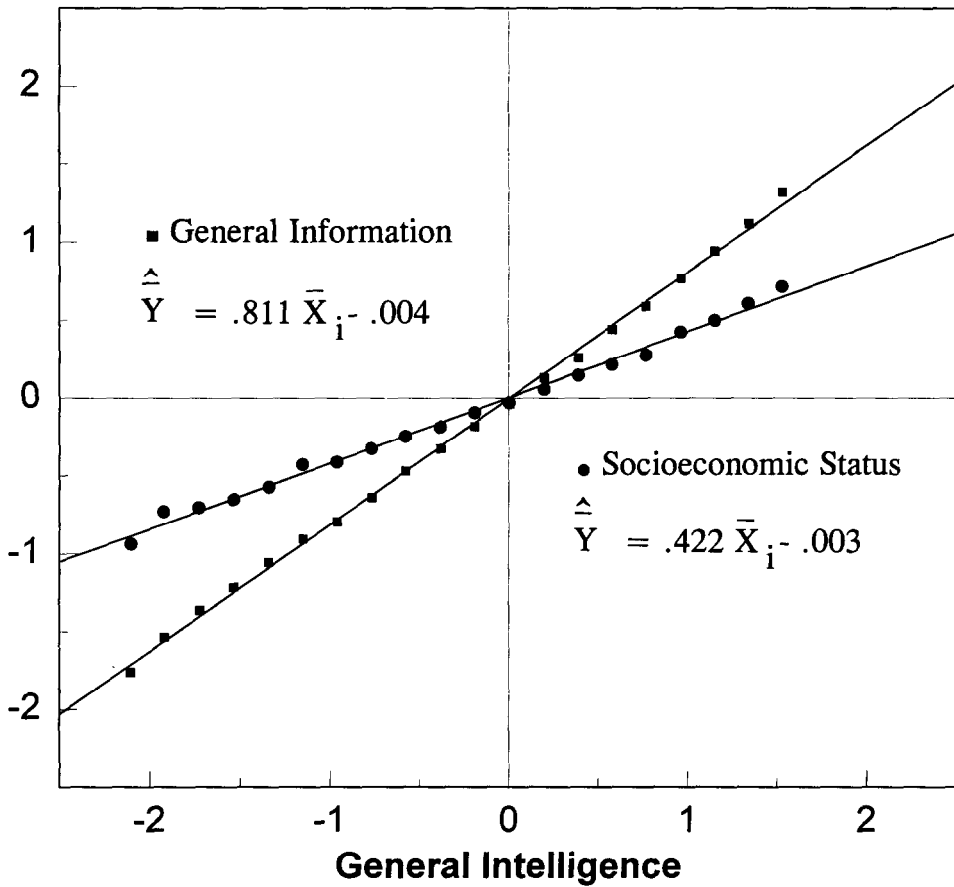
Figure 2. Plotting Bivariate Means. This sample of approximately 95,000 participants was taken from the entire Grade 12 cohort of Project Talent (Flanagan et al., 1962).

exceeded .99. This correlation is an index of the extent to which we can forecast the mean criterion from mean level of general intelligence for the series of intellectually homogeneous groups. The magnitude of the relationship among individuals between general intelligence/SES and general intelligence/general information is revealed by the slopes of the regression lines fitted to these bivariate means, .42 and .81, respectively (as opposed to the .99 correlations among means manifested by each).

To be sure, there is nothing new here in terms of uncovering statistical relationships not implicit in the familiar simple correlation that determines the slope of the line of regression. Indeed, to the extent that one's data are bivariate normal and homoscedastic (Lubinski & Humphreys, 1996), the slope of the regression line fitted to bivariate means should equal the "simple" correlation conventionally computed on the entire sample of individuals. What this graphic presentation of

bivariate means does do, however, is to make data more interpretable and harder
to ignore. It highlights the magnitude of rate of change across successive groups
and the precision with which we can predict a group's criterion mean from its
mean intelligence. Within each group, the average difference between people in
terms of their general intelligence is $< .10$ SD.

Substantively, the example presented here highlights the generality of the con-
struct of general intelligence (as discussed earlier concerning data from Table 1).
Notice, as one moves across successive intellectual gradations, the markedly dif-
ferent rates of change between general information and SES. Anyone holding the
position that conventional measures of general intelligence primarily index nar-
row academic skills and/or privilege associated with high SES is invited to ex-
plain the covariation between the highly general competencies associated with
general intelligence (assessed primarily through arithmetic reasoning, abstract
reasoning, and reading comprehension) and the narrow bits of specialized knowl-
edge associated with General Information (i.e., about art, law, hunting, fishing,
the Bible, medicine and health, engineering, photography, architecture, journal-
ism, foreign travel, military, accounting, business, sales, practical knowledge,
clerical information, colors, outdoor activities, etiquette, sedentary games, the-
ater and ballet, and food). What is especially interesting here, in view of the
widespread misconception that measures of general intelligence assess narrow
academic skills associated with privileged environments, is that the covariation
(rate of change, or slope) between general intelligence and general information is
approximately twice that between general intelligence and SES. Acquiring many
different kinds of esoteric bits of cultural knowledge is often associated with being
raised in highly privileged homes, but the covariation (slope) between general
information and SES is relatively modest in comparison, and essentially equiva-
lent to that of general intelligence and SES, namely, $r = $ slope $= .45$.

This example illustrates how examining group trends (graphs of bivariate means)
can help answer questions about competing causal explanations. Namely, is it SES
or general intelligence that serves primarily to determine the amount of cultural
information a child learns? If high levels of environmental privilege (SES) are
thought to be responsible for performance on general intelligence tests, then why is
it that groups occupying extraordinary levels of SES, say, the top 0.5%, evince only
relatively modest levels of general intelligence and general information about
their culture (approximately 1 SD above the norm)? And why is it that groups of
people at extraordinary levels of general intelligence, again, top 0.5%, by and
large reside in relatively modest homes (approximately 1 SD above the mean in
SES), and yet they were able to acquire impressive amounts of cultural knowledge
(about 2.0 SD above the mean in general information). Clearly, exceptional
homes are not required to learn superior amounts of cultural knowledge or to
develop sophisticated levels of intellectual functioning. An exceptionally privi-

leged home far from guarantees exceptional intellectual development or, by the same degree, learning an impressive amount of cultural information.

Of course, much is already known about the relation between intelligence and SES. We know from meta-analytic reviews of twin and adoption studies, for example, that the correlation of IQs between biologically unrelated children reared together as siblings approaches zero by adulthood (Bouchard, 1996; Plomin, 1994; Rowe, 1994; Scarr, 1992). So variation in childhood SES (as conventionally measured) appears to exert little influence on individual differences in general intelligence manifested at adulthood. Nevertheless, this illustration serves to show how our graphic approach, plotting and correlating means, puts the theoretically meaningful implications of multivariate information in a clearer light for policymakers and the public. It forces you to look at the data; it does not obstruct it so much. Such clarity is not found by simply reporting values of zero-order, partial, or multiple correlations.

Recognizing the Utility of Modest Individual-Level Correlations and Regression Slopes

As our reference to smoking behavior indicated, predictive validity correlations should not be lightly dismissed simply because they are small (especially those that point to multiple direct effects, all in the same direction: e.g., smoking, unsafe sex, sharing needles). There are at least three relevant considerations when evaluating small correlations (or modest regression slopes): base rates, effect sizes, and cascade effects of the phenomena of interest. One hears and sees nowadays much discourse surrounding certain "small" correlations with general intelligence. All too frequently, furthermore, correlations are inappropriately squared for calibrating their potential utility (cf. Gordon, 1997; Gottfredson, 1997b; Lubinski & Humphreys, 1996; Rosenthal, 1990). However, the implications of small correlations are often greater than their magnitude, squared or unsquared, would suggest.

Base Rates. One of the best ways to demonstrate the utility of correlations of modest magnitude is through the Taylor-Russell (1939) tables, which require three bits of information: the base rate, the selection ratio, and a correlation. When the base rate of a phenomenon for a given population is relatively infrequent, say, the likelihood of success in a training program = 30%, and the correlation between a test score and training success (e.g., *success* = 1, *failure* = 0) is .30, administering the test to samples from this population and selecting the top 20% will secure 46% successful employees. Thus, the test contributes 16% incremental validity (i.e., hit rate minus base rate, or 46% − 30% = 16%) over the base rate, thus reducing the percentage of *unsuccessful* trainees from 70% to 54%. In employment settings, when training is expensive and time consuming, and failures are costly, consulting a test with "modest" validity can result in huge savings. Yet,

following a *Newsweek* story that cited a National Research Council report (Wigdor & Garner, 1982) documenting that cognitive tests were unbiased for predicting performance across multiple demographic groups, the president of the National Academy of Sciences wrote in a response to *Newsweek*: "The prediction for any group is not strong—about nine percent of the variation in job performance" (Alberts, 1994, p. 22). Saying that a test accounts for only 9% of the variation in training success or performance (.30 squared) is beside the point for judging its practical utility in many situations (Rosenthal, 1990).[4] This statement needs to be evaluated in the context of other information. When the likelihood of observing a particular outcome is low to begin with (as in experiencing an auto accident, developing lung cancer, passing a stringent training program, or becoming a victim of crime), and an individual or group's status on the predictor is high, small correlations add appreciable predictive power to scientific forecasts in many contexts.

Effect Size. A second context for more fully appreciating the potential utility of small correlations is found in experimental psychology. Following Cohen (1988), experimentalists routinely report *effect size* (*SD*) differences (or *d*s) between experimental and control groups to quantify the average magnitude of experimental effects. Drawing from customary practices in the field, Cohen has arbitrarily referred to *d*s of .20, .50, and .80 as small, medium, and large, respectively. There is a simple formula for converting effect size differences into correlations: $r = d \div (d^2 + 4)^{1/2}$. Using this formula, Table 2 shows that the magnitude of Cohen's small, medium, and large effect sizes is comparable to correlations of .10, .24, and .37, respectively.[5] Clearly, correlations frequently referred to as "minuscule" by many social scientists actually depict functional relationships that experimental psychologists refer to as *large*. If the effect size after 1 year of an experimental treatment concerned with improving reading comprehension were .80, the investigator could hardly be faulted for recommending application of the method. Such results would mean that students who would have been reading at the 19th percentile would often be reading close to the 50th percentile after treatment. Yet, this mean difference is comparable to a correlation of only .37! If correlations of .37 are large in the context of one type of psychological research (experimental psychology), they also should be considered large in a different

---

[4]This reasoning also obfuscates the amount of forecasting precision achievable when individuals who are relatively homogeneous on an important category or dimension are aggregated, as in Figure 2. When this is done, the standard error of estimate for the group's mean on the criterion (SEE_M) may be estimated by dividing the conventional standard error of estimate [viz., $SEE = S_y (1 - r^2)^{1/2}$] by the square root of the number of participants in each group: $SEE_M = S_y(1 - r^2)^{1/2} \div n^{1/2}$.

[5]The more general formula for the transformation exchanges $1/pq$ with 4, where p = proportion of Group 1 in combined Group-1 and Group-2 samples, and $q = 1 - p$ (i.e., proportion of Group 2s). Thus, when $p = q$, $1/pq = 4$.

TABLE 2
Effect Sizes and Their Corresponding Point
Biserial Coefficients of Correlation When Two
Treatment Groups Have Equal Ns

Effect Size	Description	Correlation
.20	Small	.100
.50	Medium	.243
.80	Large	.371
1.00		.447
1.50		.600
2.00		.707
2.50		.781
3.00		.832
3.46		.866

Note. Based on Cohen (1988, p. 22).

type of research (differential psychology). If a correlation of .37 represents a large and important change on the criterion that is attributable to treatment, logic would suggest that it also represents a large and important difference or "effect" in ("untreated") populations. Whether or not data are used to design practical applications involves a value judgment, but decision makers should not be deprived of data simply because of misconceptions about the proper size of r.

Cascade Effects. When statistical trends reflect functional relationships with direct and indirect implications for large segments of our society, as in the costs of smoking and passive smoking, informed policy decisions may very well engender a cascade of positive effects, such as reductions in work absenteeism, fewer health-related insurance claims, and higher economic productivity, that replace a cascade of negative effects. This is because the direct effect of smoking on health eventuates in a family of indirect effects with broad economic, physical, and psychological implications throughout society. The well-known direct effect indexed by the negative correlation between general intelligence and crime (Eysenck, 1970; Gordon, 1987; Wilson & Herrnstein, 1985) also produces broad bands of indirect effects with respect to victims, property damage, insurance rates, and law enforcement facilities and personnel. Small correlations can thus have wide-ranging implications, but unless multiple outcome criteria are assessed, those implications will remain undocumented.

Using Individual Differences Data to Forestall Spurious Causal Interpretations of Obvious Group Statuses: Exemplified by Race and Gender

Competing interpretations of data are a fact of life, but consumers of epidemiological and social science findings must be particularly alert, because the reliance

on conspicuous but often superficial demographic groupings is especially likely to give rise to spurious inferences. This is highly likely when relevant individual differences measures are omitted from the inquiry. For example, significant correlations between gender, race, ethnicity, and SES, on the one hand, and measures of social phenomena, on the other hand, are commonly observed in epidemiological (Friedman, 1994) as well as in psychological research (Tyler, 1965; Willerman, 1979). Those correlations are often assumed (mistakenly) to represent causal effects of gender, race, ethnicity, and SES per se. It is our hypothesis, however, for which there is a good deal of empirical support, that measures of relevant individual differences will consistently produce higher correlations with social phenomena than will traditional demographic groupings (Herrnstein & Murray, 1994). A correlation with demographic group membership may be highly misleading, when viewed in isolation, and thus lead to inappropriate causal inferences. The cognitive domain provides an easily generalizable example of the importance of going beyond such correlations with demographic categories.

We begin our example by noting that it is sometimes important to conduct a more refined analysis of the cognitive domain than general intelligence alone affords (Humphreys et al., 1993a; Lubinski & Dawis, 1992). Two important dimensions that add incremental validity beyond general intelligence to the prediction of performance in a variety of educational and vocational settings are Vernon's (1947, 1961) verbal–numerical–educational (v:ed) and spatial–mechanical–practical (k:m) major group factors. Females, as a group, display higher means on v:ed, whereas males are higher on k:m. Consideration of these two major ability factors reveals how correlations with gender may be misleading.

Many large-scale studies have shown that gender (with male = 0, female = 1) is predictive of both clerical (positively) and mechanical–technical (negatively) performance and occupational choice (Humphreys, 1962, 1986; Vernon, 1947, 1961). By contrast, a conventional measure of general intelligence does not inform us as to whether anyone is likely to perform more competently in clerical versus mechanical–technical domains. On the other hand, measures of v:ed and k:m, respectively, are capable of doing so. And they will be more predictive of clerical and mechanical performances than predictions based on gender. That is, v:ed and k:m measures will absorb the forecasting efficiency offered by gender and add incremental validity to boot. It would be contrary to an employer's self-interest (as well as now being illegal) to use gender as an explicit criterion for selection for engineering or office management positions. Gender is only a crude surrogate for v:ed and k:m. It is not group status, per se, but the individual-differences measures that more precisely assess skills relevant to performance within these two broad occupational arenas.

Similarly, dummy-coding race (*Black* = 0, *White* = 1) results in predictive validity for a variety of educational and vocational performances (Humphreys, 1988; Hunter, Schmidt, & Hunter, 1979; Schmidt, 1988). Nevertheless, general

intelligence typically subsumes this relationship and adds incremental validity beyond it. The point is that when investigators have available construct-valid measures of general and more circumscribed abilities (viz., general intelligence vs. v:ed and k:m, respectively) relevant to general and more circumscribed performances (viz., overall vs. clerical and mechanical–technical, respectively), holding such measures constant by computing partial correlations between demographic groups and job performance typically shrinks the group correlations to zero. But the inverse is not true. Namely, holding group (gender, race) constant still leaves a significant amount of overlap between relevant individual-differences measures and school or work performance. Whatever the causes of group differences in social phenomena, measures of individual differences typically reflect those causes more effectively than does membership in demographic groups.

Using Constellations of Individual Differences to Better Understand Sources of Group Differences: Gender Differences in Majors and Career Choice. As mentioned earlier, it is frequently profitable to combine ability assessment with other relatively distinct individual-differences dimensions. In vocational psychology, abilities and preferences are frequently examined conjointly (Dawis, 1992; Dawis & Lofquist, 1984; Williamson, 1965) for ascertaining career paths that individuals are likely to find satisfying (fulfilling) and within which their performances are likely to be found satisfactory (competent). "Interests, values, and preferences are stable dispositions . . . that have proven their worth in predicting (a) occupational membership, tenure, and change, (b) job and career satisfaction, and (c) worker satisfactoriness, though secondary to abilities in this regard" (Dawis, 1991, p. 833).

In much contemporary social science literature on women and engineering, group differences on key individual-differences variables are ignored. Large male/female ratios in graduate training programs are pointed to as cause for concern. Yet, gender differences in means *and* variances on three-dimensional spatial visualization and mathematical reasoning collectively serve to generate disparate male/female proportions at the upper tails of these ability distributions (Hedges & Nowell, 1995; Lubinski & Dawis, 1992; Stanley et al., 1992). These are the two abilities most critical for engineering, regardless of gender. This helps to explain large male/female ratios. Perhaps more importantly, however, when preferences are consulted, pronounced gender differences are observed in vocational interests for working with people versus things and in pursuing humanistic versus scientific disciplines; females tend to gravitate toward the former, males toward the latter (Lubinski & Benbow, 1992; Humphreys, Lubinski, & Yao, 1993). These gender differences in preferences are routinely observed in mathematically gifted females who possess more quantitative reasoning ability than the typical physical scientist (Achter, Lubinski, & Benbow, 1996, p. 76; Lubinski, Benbow & Ryan, 1995; Lubinski & Humphreys, 1990a; Lubinski, Schmidt, & Benbow, 1996). Giv-

en these systematic individual differences in personal attributes critical for chan-
neling educational and vocational development down different paths, differential
male/female proportions in certain careers is not really a mystery. They are, at
least partly, forecastable and understandable (Humphreys & Lubinski, 1996).

There is a tendency for mathematically talented females (who more often than
not are verbally exceptional too, because of the positive manifold running through
cognitive abilities) to prefer to secure educational credentials and excel in busi-
ness, the humanities, law, management, medicine, and the biological and social
sciences, rather than in subject matters consisting primarily of inorganic scientific
content (Lubinski & Benbow, 1992, 1994; Lubinski, Benbow, & Sanders, 1993).
Yet, yearly, pages of journal space are given to investigators who neglect these
important individual differences. But it is unlikely that doing so is a favor to
intellectually gifted females who prefer to develop their talents in other areas, or
to engineering faculty who are trying conscientiously to increase their enrollment
of women and who continue to be compared unfavorably with faculties in medi-
cine and law. Should we really expect graduate engineering programs to reflect
male/female ratios comparable to medicine and law? If not, should individual
differences dimensions be consulted to generate male/female expectations? Is
there a more appropriate basis for generating expectations, which also captures
the familiar phrase in counseling psychology, "Be true to yourself?" Surely col-
lecting data from gifted females regarding what they prefer to do, reporting their
responses, and factoring their preferences into scientific interpretations of gender
differences is desirable.

Using Polychotomous Demographic Groups to Forestall False
Inferences About the Source of a Group Difference: Exemplified by
Black–White Differences in Low Birth Weight

Thinking about polychotomous groups simultaneously can be highly informative.
Actually, doing so can be helpful in uncovering inconspicuous causes, as well as
in exposing certain presuppositions, which, in reality, are unraised empirical
questions in need of empirical answers. Consider the following analysis of low
birth weight and infant mortality, which appears to shed light on a prevailing issue
in epidemiological research.

Relevant data for our point are available from the Monthly Vital Statistics
Report series of the Centers for Disease Control and Prevention/National Center
for Health Statistics. In these reports, groups are defined on attributes such as age,
education, smoking habits, sex, ethnicity, and race. As is true for most nonexperi-
mental data obtained without control over independent variables and random as-
signment to groups, causal relations cannot be inferred directly from the data in
these reports. The data do provide guidance, however, for developing hypotheses
that direct research toward plausible causal paths. The other side of this coin is
that the absence of covariation between variables that are purportedly causally

related serves to question the plausibility of certain hypotheses (recall the near-zero correlation at adulthood between IQs of biologically unrelated children reared together).

Every year, at time of publication of annual reports on births and deaths in this series, selected information is released to reporters at press conferences informing citizens about the incidence of low birth weight and infant death rates in our population. The incidence of both is higher among Blacks than among Whites. The media attribute each of the two racial differences to differences in medical care available to mothers in the two populations. News stories typically point out that this country lags far behind other highly developed countries in our greater incidence of these health problems. Commentators readily conclude that the data constitute evidence of continuing discrimination in the United States, understandably a reasonable hypothesis. But these conclusions—like many such produced by social science—neglect pertinent data. Hypotheses are more likely to be valid when they take account of all promising aspects of the evidence concerning the phenomenon in question.

The monthly series of publications, including the specific reports in which comparisons are made between Blacks and Whites, contain corresponding information about other ethnic groups. However, this additional information—available for more than 10 years—is routinely ignored by scientists, journalists, and commentators alike. If considered, this added information indicates that it is highly unlikely that medical care or general economic privilege is the cause of Black–White *differences* in low birth weight and infant mortality.

Consider, for example, the incidence of low birth weight in 1984 among White, Black, American Indian, and Alaska native mothers (National Center for Health Statistics, 1987). Table 3 contains data from this publication showing that mothers in the three minorities are very similar to each other in characteristics associated with the medical care hypothesis (e.g., no prenatal care, number of prenatal visits). However, mothers in two of the three minorities, American Indians and Alaska Natives, are much more similar to White than to Black mothers in incidence of low birth weight. The three minority groups differ from the White mothers on all measures of medical privilege, and yet two of them are more similar to Whites than to Blacks in incidence of low birth weight babies (viz., White = 5.6%, Black = 12.3%, American Indian = 6.1% and Alaskan native = 5.9%).

There is no denying that availability of medical care has an effect on low birth weight, but it is questionable that medical care accounts for the difference between Blacks and Whites because it seems to have no parallel effect on rates for American Indians and Alaskan natives. The incidence of low birth weight has dropped steadily for a number of years in minority groups as well as in the White majority. The quality of prenatal care is likely to be one of the causes, but that cause appears to make a trivial contribution to the Black–White difference.

This conclusion is reinforced by recasting the information in Table 3 into the

TABLE 3
White, Black, American Indian, and Alaskan Native Mothers With Selected
Characteristics, 1984 (%)

Characteristics	White[a]	Black	American Indian	Alaskan Native
Age (years)				
<20	11.1	23.7	20.2	17.7
20–30	63.8	59.1	62.0	62.3
>30	24.9	17.3	17.8	20.2
Actual number of out-of-wedlock births/1000 live births	141.7	530.2	363.2	374.9
Education (years)				
<12	18.0	33.1	38.7	30.7
12	43.1	42.8	41.4	49.0
>12	38.9	24.0	20.0	20.4
Fetal gestation (weeks)				
<37	7.8	16.9	10.9	12.4
37–39	36.7	40.7	36.0	40.6
>39	55.4	42.6	53.1	47.2
Prenatal care started at (months)				
1–3	79.6	62.2	59.6	65.5
4–6	15.7	28.2	27.8	25.5
7–9	3.3	6.4	9.3	7.1
No prenatal care	1.3	3.3	3.3	1.9
Actual number of prenatal visits	12.0	10.3	10.0	9.8
Low birth weight infant	5.6	12.3	6.1	5.9

[a]Includes Hispanics.

Table 4 format. Percentages of low-birth-weight infants from the same report replace the percentages of the mothers in each of the characteristics listed. Low birth weight among Blacks is not associated differentially in any obvious fashion with any of the characteristics studied. Percentages for Blacks are highest in every row (e.g., for mothers less than 20 years of age: White = 7.7%, Black = 13.6%, American Indian = 6.5% and Alaskan native = 5.9%; for mothers with less than 12 years of education: White = 8.4%, Black = 14.5%, American Indian = 6.9% and Alaskan native = 6.4%). There is clearly a missing variable—something critical that has not been measured. There is a "neglected aspect."

Table 5 extends the number of minorities that can be compared to Blacks and Whites to two Hispanic groups that are approximately the equivalent of Blacks in education and economic deprivation and to two Asian American groups comparable to Whites in economic privilege. The data in Table 5 are also more recent (National Center for Health Statistics, 1994a). Mexican Americans are similar to Blacks on various measures of prenatal care (e.g., late or no prenatal care, smoker,

TABLE 4
Low-Birth-Weight Infants Among Mothers Categorized in Table 3 (%)

Characteristics	White[a]	Black	American Indian	Alaskan Native
Age (years)				
<20	7.7	13.6	6.5	5.9
>30	5.3	12.0	6.6	7.9
Marital status				
Unmarried	8.6	14.0	6.9	6.8
Married	5.1	10.0	5.7	5.3
Education (years)				
<12	8.4	14.5	6.9	6.4
>12	4.4	10.0	4.7	5.1
Fetal gestation (weeks)				
<37	39.6	42.6	32.1	29.0
37–39	4.1	7.8	4.3	3.0
>39	1.5	4.0	2.7	2.8
Prenatal care started at (months)				
1–3	5.1	11.5	5.7	4.4
7–9 or no prenatal care	8.8	16.3	7.7	11.6

Note. Categories do not precisely mirror those in Table 3 because the report did not make that possible.
[a]Includes Hispanics.

drinker) but they are, nevertheless, more like Whites, Chinese Americans, and Japanese Americans in their incidence of low birth weight (White = 5.7%, Black = 13.4%, American Indian = 6.2%, Mexican American = 5.6%, Chinese American = 5.0%, and Japanese American = 7.0%). The percentage of low birth weight among Puerto Ricans (9.2%) is revealing, because Puerto Ricans are intermediate between Blacks and Mexican Americans in cultural-genetic mix, and their percentage of low birth weights is intermediate between these two groups.

Table 5 includes infant death rates for Whites, Blacks, and the two Hispanic groups. These data were not available for American Indians and the two groups of Asian Americans in this report (National Center for Health Statistics, 1994b). The rates for Whites and Mexican Americans are approximately equal, both are lower than the rate for Blacks, and that for Puerto Ricans is again somewhat intermediate.

The same report breaks down the causes of infant death into the 10 most frequent categories and one "all other" category, but only for Whites and Blacks. These data appear in Table 6. We have added Black/White ratios. The smallest ratio (for congenital anomalies) and the largest (for short gestation and low birth weight) differ from each other substantially: 1.25 versus 4.99. Otherwise, the homogeneity of the ratios is impressive, the variation being entirely between 2.1 and 2.9. Whatever the missing variable(s) is (are), it appears to have much the same effect on all but two of these many different factors in infant death.

TABLE 5
Births Associated With Selected Characteristics of Mothers and Infants in Seven Ethnic
Groups (%)

	White	Black	American Indian[a]	Mexican American	Puerto Rican	Chinese American	Japanese American
Mother							
Early prenatal care	84.9	64.0	62.1	62.1	67.8	83.8	88.2
Late or no prenatal care	2.8	9.8	11.0	10.5	8.0	2.9	2.4
Smoker	19.7	13.8	22.5	4.3	12.7	1.7	6.6
Drinker	2.5	3.3	6.6	1.0	2.6	1.2	1.5
Weight gain < 16 lb	7.9	15.9	14.0	13.0	11.3	7.0	9.3
Infant							
Preterm	8.7	18.5	11.6	10.4	13.2	7.0	7.9
Low birth weight	5.7	13.4	6.2	5.6	9.2	5.0	7.0
High birth weight	12.7	5.3	12.3	9.7	7.1	6.2	5.1
5-min Apgar >7	1.2	2.6	1.4	1.3	1.4	.9	.8
Infant deaths[b]							
<28 days	4.2	10.7		4.2	5.2		
≥28 days, <11 months	2.6	6.0		2.6	2.6		

[a]Includes Aleuts and Eskimos.
[b]Per 1000 live births.

Consistent with an earlier statement about correlations not under experimental control, the data in these four tables do not pinpoint a cause. Nevertheless, for Blacks, there is more to their inordinate numbers of low birth weights and infant death rate than differences in objective indices of prenatal, perinatal, and postnatal medical care can explain. The SES level of the mother, a more general variable than prenatal medical care, does not distinguish among the racial or ethnic groups in these data in the way required to match the differences in pathologies. Means on a standard test of general intelligence are more plausible, but some problems are encountered in matching the pattern of group differences in general intelligence to group differences in the observed pathologies. Yet, given the pattern of findings, a hypothesis does emerge that cannot be ruled out, which involves a systematic source of intellectual differences.

A major group factor, spatial–mechanical–practical (k:m) abilities—an important component of general intelligence—fits the patterns observed in Tables 3, 4, and 5 (and does so more accurately than general intelligence). Among Black–White differences in cognitive functioning, the smaller differences are in verbal abilities, the larger differences are in nonverbal reasoning, especially the spatial-visualization ability (Jensen, 1980). These latter abilities are more closely related to the Performance IQ than the Verbal IQ on the Wechsler scales. The additional five groups in Table 5 are more in line with Whites on these spatial-visualiza-

TABLE 6
Mortality Rates[a] for the 10 Leading Causes of Infant Death Under 1 Year, Plus Black/White Ratios

	White[b]	Black	Black/White
Congenital anomalies	177.0	219.3	1.25
Sudden infant death	101.2	218.4	2.16
Disorders from short gestation and low birth weight .	60.2	300.6	4.99
Respiratory distress	41.3	104.7	2.54
Affected by maternal complications	30.1	69.2	2.30
Complications from placenta, cord, membranes	20.6	44.2	2.15
Infections specific to prenatal difficulties	17.9	46.4	2.59
Accidents	16.6	37.3	2.25
Intrauterine hypoxia and birth asphyxia	12.6	27.5	2.18
Pneumonia and influenza	11.2	32.1	2.87
All other causes	203.7	586.1	2.88

[a]Per 100,000 live births.
[b]Includes Hispanics.

tion/mechanical-reasoning abilities relative to Blacks—with the Asian Americans perhaps being a bit superior (Lynn, 1987). An ancillary finding, based on a large adolescent sample (Lubinski & Humphreys, 1992), suggests that (nonverbal) quantitative reasoning ability is a more salient covariate of physical well-being than are traditional measures of SES. But the hypothesis that a major group factor of cognitive functioning may partly explain the findings observed in Tables 3 through 5 is only one of a host of possibilities. For example, are there temperamental dimensions, cultural differences in diet, substance abuse, or anatomical or physiological differences that fit the pattern required? To answer these questions, as well as others of equal importance, we need to obtain and report in scientific journals and to the public dependable data concerning these issues.

Finally, to further highlight the long-standing bias on the part of researchers *not* to analyze data on race and covariates thereof in more comprehensive detail note the following. An analysis quite similar to the above was reported in *Science* almost 30 years ago in Kuttner and Lorincz's (1968) article, "Utilization of accentuated environmental inequalities in research on racial differences." Kuttner and Lorincz (1968) questioned whether Black–White *differences* in educational achievement uncovered in the well-known Coleman Report (Coleman et al., 1966) could be solely attributable to differences in environmental deprivation, given other data contained in the same study. American Indians outperformed Blacks on scholastic tests despite suffering more intensely from environmental deprivation (on the same measures used to document the black–white differences in environmental deprivation and privilege). Coleman et al. (1966) initially collected all the necessary data to reveal this significant quasi-experimental trait-by-treatment interaction, yet this important fact went unmentioned in the 1966 re-

port. Black–White differences occur on variables that happen to be conspicuous (like differences in age, occupation, residence, and social class). And these differences do, indeed, have predictive validity for many important phenomena relevant to social policy. However, as indicated earlier, conspicuous demographic differences are often superficial with respect to achieving more comprehensive, causal understandings of human phenomena. Frequently, individual-differences dimensions absorb the forecasting efficiency offered by such group membership categories and add incremental validity beyond them. In so doing, individual-differences dimensions relegate conspicuous group membership categorizations—like race or gender—to the status of an unlikely causal determinant. Under these circumstances, the individual-differences dimensions emerge as *plausible,* but not unquestionable, candidates for achieving causal status themselves.

In many other such large-scale data banks, like Coleman et al.'s (1966), we suspect that data exist for similar kinds of reanalyses. We encourage readers to embark upon them. The data merely need to be categorized and analyzed more systematically. Unfortunately, however, this is not always easily achieved. Sometimes huge studies are executed that result in publications and data banks making reanalysis of the sort conducted here difficult. Whatever future investigators do when building large-scale probability samples, they should code and present data in ways that afford respecifying groups nested within multiple configurations of demographic and individual-differences variables.

Hazards of Ignoring Individual Differences: Possible Iatrogenic Effects in Education and Race Relations

Thus far, we have concentrated our concerns on problems that are more likely at lower levels of general intelligence. Is it profitable to conduct parallel analyses at the upper end of the general intelligence continuum as well to reveal possible maladaptations and indirect effects harmful to oneself and others? We think it is. The following example is once again useful for illustrating how conventional demographic categories and individual-differences dimensions can complement one another. Barbara Lerner (1991, 1995) has used the term iatrogenic effects (illnesses or deformities caused by treatments) to characterize some current educational practices. Lerner (1991, 1995) argued that relaxing standards for educational opportunities and credentials with the aim of providing members of underrepresented groups more of both has resulted in *less* learning overall. If Lerner is right in her analysis, and there is a good chance that she is, some current practices in our colleges and universities also may foster socially disruptive, iatrogenic effects when they ignore racial disproportions in general intelligence. By placing in competition readily identifiable groups, whose average levels of general intelligence and, accordingly, whose average levels of academic achievement are con-

spicuously different, well-intended policies can stimulate iatrogenic effects both directly and indirectly.

Seligman (1992) and Sowell (1992) have underscored the impressive amount of apparent interracial harmony in the military today, an organization where racial quotas are not exercised, and a premium is placed on merit. Humphreys (1991) has commented similarly. White and Black applicants for enlistment and for assignment to specialities meet the same test-score requirements. (One commonly hypothesized indirect effect of this policy is the lack of racial segregation in military messhalls, see also Lerner, 1995, p. 57.) By contrast, consider Figure 3. It reveals the mean group differences in admittees' intellectual aptitudes on numerous U.S. college campuses.

Sarich (1990) has shown that at Berkeley the undergraduate student body has become two-tiered. In recent years, there is only a 10% overlap between the SAT distribution of the Asian and White undergraduates, on the one hand, and the distribution of the Hispanic and Black undergraduates, on the other. Further, very few Hispanic or Black students score beyond the *average* for Asian or White students, which makes it exceedingly difficult for the former to distinguish themselves academically (cf. Jensen, 1991). Readers are asked to pause briefly, at this point, and ask themselves to project how differences of Figure 3's magnitude might manifest themselves if they were evinced by females over males or males over females on the college campus or high school they attended. A hypothesis that emerges is that iatrogenic effects are likely to occur whenever pronounced ability differences co-occur with a conspicuous group difference and both groups are required to compete with one another in a particular setting.

Sowell made the following argument:

> Dogmatists have attributed the high attrition rates of minority students to racism on white campuses, rather than to the mismatching indicated by test scores. However, if one goes beyond dogmas to evidence, the role of supposedly "irrelevant" academic criteria becomes clear: While only 22% of the Hispanic students preferentially admitted to Berkeley had graduated five years later, more than half of the Hispanic students admitted under normal academic standards had graduated. Figures for black students were similar. If the all-purpose explanation is racism, then why did this racism have such radically different effects on people of the same race with different test scores?
>
>   As Professor Clyde Summers predicted long ago, this mismatching problem has not been confined to the top echelon schools. As each tier finds its normal pool of minority students pre-empted by a higher tier, it must in turn pre-empt the minority students who would normally qualify for colleges in a lower tier. (Sowell, 1992, p. 146)

According to Sowell (1992) and others (Cohen, 1996; Humphreys, 1988; Stanley, 1971), this type of mismatching, beginning with the most selective institutions where there is the most extreme shortage of minority students, and proceeding downward, augments the number of Black students who fail to graduate, because
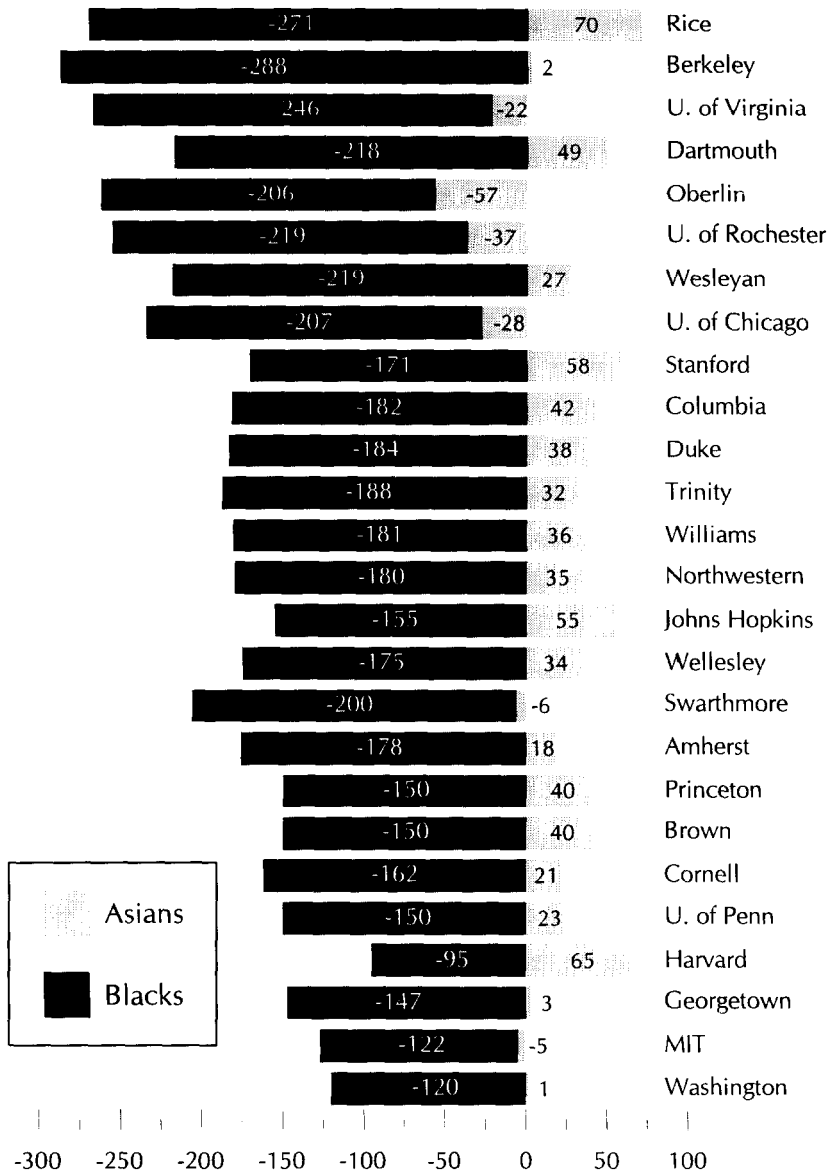
Figure 3. SAT - point difference from the white mean. Reproduced from Herrnstein and Murray (1994, p. 452) by permission of the publisher.

the performance requirements happen to be set just beyond the academic readi-
ness of many minority students within each tier. It is not implausible that this
systematic racial differential in academic readiness at each and every tier en-
genders interracial tension. (Social psychological inferences from attribution the-
ory might be a useful mechanism for conceptualizing why groups with academic
readiness deficits, for the learning environments they are in, look outside them-
selves for explanations of their shortfall in achievement. The likelihood of attribu-
tion mechanisms operating in this context is enhanced if institutions tell students
at the outset that they are all comparable in capability.) Race may be seen as a
causal antecedent because it is so conspicuous, but it may not be the underlying
determinant for group differences in achievement.

Perhaps it is not the color of one's skin that determines whether a given college
environment seems hostile or supportive, but rather, the degree to which students
are able to meet the learning demands placed on them. There is much in educa-
tional research to support the latter proposition (Becker & Carnine, 1980; Benbow
& Stanley, 1996; Englemann & Carnine, 1982), and we should seek more data as
well. Sowell (1992, p. 144) reinforced this idea by noting that Black students at
MIT score in the top 10% nationwide on the mathematics portion of the SAT,
which is quite impressive. Nevertheless, this degree of quantitative reasoning
ability falls at the bottom 10% of MIT students, because MIT is a world-premiere
university. Is this a problem? Sowell thought so. Black students at MIT have an
inordinate dropout rate—25% fail to graduate, and those who do earn signifi-
cantly lower grades than their classmates, in spite of the fact that most of these
students could earn degrees with distinction at a number of highly regarded, but
lower-tier, institutions.[6] Like Sowell (1992), we can only speculate on the degree
of personal hardship experienced by these exceptional students and the loss of
talent when they drop out of the math/science pipeline. If universities more readi-
ly provided information like the kind collected by Sowell, more effective steps
might be taken to enhance the likelihood of talented Blacks securing the advanced
educational credentials commensurate with their abilities. We should do whatever
we can to ensure that everyone has the opportunity to secure educational creden-
tials commensurate with their abilities (cf. Benbow & Stanley, 1996; Lubinski,
1996). To do so, we must acknowledge the possibility that some current policies,
although well intended, may, following Lerner (1991, 1995), foster iatrogenic

---

[6]When presented with this problem, Robyn M. Dawes (personal communication, 26 April 1996)
suggested that students be provided with expectancy-tables—forecasting the likelihood of gradua-
tion, and GPA—across the full range of test scores. This would enable students to make more
informed choices. We agree. As a matter of fact, in a recent review of the contribution of individual
differences research to counseling psychology, Dawis (1992) noted that expectancy tables are a
neglected aspect in many contemporary tests manuals and counseling settings. Years ago, such
tables were standard tools found in all well-equipped counseling centers. Dawis (1992, p. 15)
recommended that they still should be.

effects. A "neglected aspect" does not cease to operate when we choose to ignore it.

Once again, as in most epidemiological and social science evaluations of social policy, it is wise to collect as much information as possible on the multiple direct and indirect effects of a social problem. For example, it is well known that selection criteria are frequently relaxed to achieve more representative proportions of minorities in schools of medicine and law (Cohen, 1996; Davis, 1976, 1986; Gordon, 1988; Herrnstein & Murray, 1994; Sowell, 1992). The rationale that this offsets past discrimination and other unenlightened policies and past practices is offered as justification. These, indeed, are relevant considerations to weigh. Other relevant variables for medical schools to examine are offered by Bernard Davis (1986), late professor emeritus at Harvard Medical School, based on his experience at Harvard in the mid-1970s. He showed how weakening entrance standards to achieve racial balance enhances the likelihood of decreasing the standards for graduation and professional competence, in particular by decreasing the number and intensity of basic science courses, changing from letter to pass–fail grading, allowing students to retake critical exams several times, then allowing students to retake critical exams composed of the same questions, and, finally, dispensing with past practices of publicly reporting Harvard's standing on national medical exams required for certification (see Davis, 1986, pp. 171–191). From a public policy point of view, such decisions could have implications for advances in medical science, optimal patient care, and the number of malpractice cases. They could foster broad-spectrum iatrogenic effects as a function of the extent to which standards are relaxed. Again, such data can never be used to dictate specific policies, because they do not speak to our goals; but the aforementioned criteria are relevant to weigh in arriving at policy decisions and should be collected routinely and made available to the public.

DISCUSSION

This article stressed the importance of assessing general intelligence and other dimensions of individual differences in epidemiological and social science research. We discussed the many ways in which general intelligence may be assessed, illustrated a method for putting in clear light individual-differences data for policy research and development, and tried to stress throughout (as well as in what follows) that all of our proposals are offered as *complements,* not replacements, to preexisting methods and tools. In our view, adding general intelligence to many ongoing and future epidemiological and social science investigations will serve to illuminate the etiology of maladies as well as factors relevant to creating and executing optimal interventions and preventions. It will also contribute to clearing up some ostensibly mysterious findings on group differences, by provid-

ing a conceptual framework for making them not only forecastable, but more understandable.

In a way, our proposal to incorporate differential psychology into epidemiological research is not new. Approximately 30 years ago, Morris (1967) argued that the popular teaching device, the "epidemiological triangle" (agent, environment, and host) is too simplistic. He stressed the importance of the behavior of individuals and groups for modifying the effects of health-related intrinsic (host) and extrinsic (environment) factors. Subsequently, Stewart expanded upon this position:

> 1. The probabilities of coming into contact with most infectious agents are a function of the behaviors of an individual or a group. 2. Host resistance is left undeveloped, acquired, or even exhausted by previous behaviors. Host resistance is also found to be correlated with certain measurable psychological traits. 3. The environment of modern urban man has been importantly shaped by human behavior . . . and culture has without doubt altered environment drastically—in some ways promoting man's health and in other ways undermining it. The intrusion of way of life, behaviors, and emotions into the development of chronic noninfectious diseases is even better documented. [He concludes] Inasmuch as psychology is one field of scientific study of human behavior, it follows from the above that psychology has a great potential contribution to make to the scientific study of health and disease in human communities—that is, to modern epidemiology. (Stewart, 1972, p. 140)

Today, indeed, health psychology is a large domain within psychological science. (It even has its own American Psychological Association journal.) Yet, systematic sources of individual differences involving intellectual abilities remain absent from most health psychology treatments (Adler et al., 1994; Adler & Matthews, 1994; Anderson & Armstead, 1995), which tend to focus more on SES. Stewart's (1972) volume anticipated certain aspects of our approach, by discussing the potential value of incorporating personality dimensions from differential psychology into epidemiological research. Stewart provided, for example, a lengthy discussion of the Minnesota Multiphasic Personality Inventory (MMPI), though he did not touch upon the potential significance of general intelligence. Similarly, current formulations in health psychology entertain the importance of traditional dimensions of personality (Adler & Matthews, 1994), but seldom mention intellectual ability. Yet, undoubtedly, the senior author of the MMPI, Starke Hathaway, would himself have hypothesized its importance. Quite familiar to Hathaway's clinical advisees (in addition to his highly regarded clinical acumen, cf. Nichols & Marks, 1992) was his position on the central importance of general intelligence for understanding human behavior more generally. According to Hathaway, "We tend to think of general intelligence as if it only operated in educational and vocational contexts; yet, it saturates almost everything we do. It is a central component of personality" (Paul E. Meehl, 1993, personal communication). Surely Hathaway would have seen value in coupling MMPI assessments with general intellectual assessments for epidemiological inquiry and health psychology.

Indeed, the systematic examination of intellectual gradations is likely to complement many research programs. We offered examples to highlight the potential usefulness of general intelligence for clarifying issues within multiple contexts. These examples, as well as others contained in this special issue, underscore a neglected aspect in much research. They reveal a tendency for investigators automatically to turn away from general intelligence while simultaneously and often erroneously embracing familiar social science categories or dimensions as causal determinants (the favorite, of course, is SES). Simply put, testable suppositions that have not been empirically confirmed should not be accorded causal status. We all know this. But the issues we have discussed here are sensitive. Conventional demographic categories and continua are so *conspicuous,* and thinking about other *less conspicuous* factors as possibilities in emotionally detached ways can be difficult. Given this, it seems appropriate before closing to introduce two important tools from logic and the philosophy of science. To our knowledge, neither has appeared in discourse on social problems in either epidemiology or the social sciences. They merit quick assimilation in both literatures.

Carnap's Total Evidence Rule and the Fallacy of the Neglected Aspect

There is a rule in the philosophy of science called the "total evidence rule" (Carnap, 1950). It is not an abstruse or controversial theory of epistemology, but rather a guideline for inductive logic that is quite uncontested. It holds that when evaluating the verisimilitude of a theory or the plausibility of a hypothesis it is incumbent upon the evaluator to take into account all the relevant evidence when conducting appraisals. This sounds commonsensical, but the rule is frequently not observed. We believe that a sufficient evidential base has accrued to conclude that causal modeling, scale construction, experimental design, and building theories in the social sciences without regard to the considerable evidence on the wide relevance of general intelligence often constitutes a serious violation of the rule of total evidence. By violating the rule, one also commits a fallacy in logic—the fallacy of the neglected aspect (Castell, 1935)—a fallacy in reasoning stemming from omitting relevant evidence. Castell (1935, p. 33) explained the fallacy thus: "In every case of Neglected Aspect, the general character of the argument is the same: true propositions, expressing relevant facts, are heaped up; but true propositions expressing equally relevant facts are omitted." Such errors in reasoning surface in a variety of settings. Kuttner and Lorincz's (1968) reanalysis of the Coleman Report and our analysis of low-birth-weight babies exposed examples of both.

Ignoring the *possibility* that general intelligence holds causal status for a variety of critically important behaviors is no longer scientifically respectable. We must let general intelligence compete with other putative causes. We must answer questions regarding its scientific status empirically. We can no longer afford to say, "You can study that, I'll study this." It is too likely that whatever we are both

studying is a covariate of general intelligence (Gottfredson, 1997b). In the following section, some consequences of neglecting general intelligence in various research designs underscore why it is important to examine general intelligence *concurrently* with wherever else we choose to study.

Common Violations of the Total Evidence Rule

Example 1: Causal Modeling. SES is frequently cast as a causal antecedent for facilitating the development of general intellectual functioning. Thus, SES is often partialed out (i.e., statistically removed) from IQ/achievement correlates, so as to control for its presumed causal influence on IQ, which, presumably, contributes to subsequent achievement. This has been done for decades in the child development literature. Yet, as already noted, when the IQs of biologically unrelated children reared together are correlated at adulthood, they tend to correlate near zero (Bouchard, 1996; Plomin & Petrill, 1997; Scarr, 1992), indicating that shared SES has little permanent effect on individual differences in IQ. This is not to say that the environment has no effect on intellectual development, but it does appear that the environment measured by traditional SES indices (the environmental features most social scientists feel are exceedingly important) contributes little to individual differences in intellectual abilities manifested in adulthood. SES should not be partialed out in developmental designs aimed at explicating antecedents to individual differences in cognitive functioning. Doing so violates the rule of total evidence, at least when familial parent–child IQ correlations are used to support causal inferences regarding the role SES plays in engendering individual differences in intellectual development. The familial correlation between the IQs of adults and their biological parents' SES, a finding frequently observed and reported as a causal association throughout the child development literature, primarily reflects parent–child genetic similarity in IQ (Plomin, 1994; Rowe, 1994). (More extensive treatments of partialling are provided by Meehl, 1970, 1971.)

Example 2: Scale Validation. Neglected aspects can occur at more fundamental levels of psychological measurement as well. Investigators purporting to assess optimal forms of psychological functioning with innovative measures (e.g., creativity, ego development, moral reasoning) frequently launch their validation campaigns without ever considering the possibility that preexisting concepts and measures might explain their outcomes of interest more fully. Just as early measures of creativity typically explained less and seldom accounted for additional variance in performance relative to general intelligence (McNemar, 1964), contemporary measures of "moral reasoning" and "ego development" probably add little to the prediction of meaningful psychological phenomena over conventional general ability measures (Sanders, Lubinski, & Benbow, 1995). When modern measures of moral reasoning, for example, compete with markers of general intel-

ligence, they have not been shown to add incremental validity to relevant criteria of interest. Yet articles continue to be published relating "moral reasoning ability" to a host of outcome criteria, including "professional ethics" (Rest & Narvaez, 1994), with the implication that these relationships constitute more than what preexisting ability dimensions would provide—but without actually testing that supposition. Recall that differently named and constituted mental tests tend to measure the same general mental ability.

Example 3: Experimental Design. As indicated earlier, investigations aimed at uncovering the causes of maladies are only one aspect of epidemiological research. Frequently, epidemiological studies seek clues for enhancing prev⁻⁻tive measures. If such studies were to assess general intelligence, they would likely produce more efficacious interventions. For this to happen, however, all relevant information pertaining to the problem under analysis must be considered. We cannot afford to neglect any aspect. For example, a recent article in the *Journal of the American Medical Association* (Williams et al., 1995) highlighted the need to attend to patients' "health literacy." These investigators examined the health literacy of patrons at two large urban hospitals. Over 26,000 patients filled out the Test of Functional Health Literacy in Adults. The results were shocking. Large proportions of the patients were unable to comprehend simple directions for taking medication on an empty stomach, information regarding their next appointment, or the information on a consent form. Large proportions of both English-speaking and Spanish-speaking patients have inadequate or marginal functional health literacy. Results appeared to go beyond a simple literacy–illiteracy distinction, because the problem for many patients had to do with fundamental comprehension capabilities and thus, perhaps, intelligence. Preventive health care measures that do not take these data into account do more than simply violate a fundamental principle in philosophy of science. They reduce the likelihood of generating effective interventions. This idea is beginning to be communicated by some health psychologists. For example, Taylor (1991, p. 310) has recently remarked: "Factors that influence patients' ability to understand and retain information about their condition include intelligence and experience with the disorder. Some patients are not intelligent enough to understand even simple information about their case, and so even the clearest explanation falls on deaf ears." Yet, the proportion of patients that this concern is relevant to is probably underappreciated (Gottfredson, 1997b).

Just as early psychometric questionnaires (75 years ago) needed to be tailored to the sixth-grade reading level, because a large segment of the general population was not reading more complex text with comprehension, many interventions targeted toward physical and psychological well-being might profit from considering the proportion of people at risk for multiple maladies located at lower segments of the IQ distribution. Their comprehension of treatment recommendations and their understanding of preventive measures need to be assessed as well. Not doing so

violates the rule of total evidence. Where one defines this segment is arbitrary, of course, but it is useful to keep in mind that approximately 25% of the general population has IQs below 90, and approximately 10% below 80. Table 7 provides some concrete information about the vocabulary such individuals are likely to understand. If policies designed to enhance well-being are to be optimally effective within these intellectual gradations, at which persons are frequently at high risk for undesirable experiences (Humphreys, 1988; Lykken, 1995), those policies and interventions too must be tailored to individual differences in comprehension capabilities. Implementing health care interventions, once they are comprehended by high-risk populations, is difficult enough.

Few people who are familiar with the intellectual repertoires of individuals operating within the 70 to 80 IQ range (low IQ but not retarded) would find comprehension capabilities irrelevant in planning innovative interventions for this segment of the population. In fact, tailoring interventions to differential levels of general intelligence is an example par excellence of Cronbach's (1957) recommendation to combine differential and experimental psychology. It also is the most consistent example of an aptitude-by-treatment interaction in educational research (Cronbach & Snow, 1977, pp. 496–497). Unfortunately, this idea is seldom implemented outside educational settings using general intelligence, psychology's most scientifically significant source of human variation, which encompasses general problem-solving ability, acquisition of cognitive knowledge and skills, and capacity to plan.[7] Given the many behavioral contexts in which individual differences in intelligence are likely to influence outcomes, it is puzzling that this important dimension of human diversity is not consulted more often.

For a penetrating analysis of a likely mechanism for maintaining neglect of certain aspects in social science research, we recommend James S. Coleman's

---

[7]After reviewing 20 years of empirical research on trait-by-treatment interactions (ATIs), Cronbach and Snow (1977, p. 496–497) concluded: "While we see merit in the hierarchical conception of abilities, with abilities differentiated at coarse and fine levels, we have not found Guilford's subdivisions a powerful hypothesis . . . Instead of finding general abilities irrelevant to school learning, we have found nearly ubiquitous evidence that general measures predict amount learned or rate of learning or both. And, whereas we had expected specialized abilities to account for interactions, the abilities that most frequently enter into interactions are general. Even in those programs of research that started with specialized ability measures and found interactions with treatment, the data seem to warrant attributing most effects to general ability." Snow (1989, p. 22) has recently restated his position on the general factor, albeit more strongly: "Measures of general ability (G) enter interactions more frequently than other measures of aptitude, despite the fact that measures of G also typically show strong aptitude main effects. Many different measures have been used to reach this conclusion . . . Given new evidence and reconsideration of old evidence, G can indeed be interpreted as 'ability to learn' as long as it is clear that these terms refer to complex processes and skills and that a somewhat different mix of these constituents may be required in different learning tasks and settings. The old view that mental tests and learning tasks measure distinctly different abilities should be discarded, even though we still lack a theory for integrating the two."

TABLE 7
Prototypic Response Patterns on WAIS-R Vocabulary Subtest for
Three Full-Scale Adult IQs

Test	IQ ≈ 70	IQ ≈ 80	IQ ≈ 90
1. Bed	2	2	2
2. Ship	2	2	2
3. Penny	0	2	2
4. Winter	1	2	2
5. Breakfast	2	2	2
6. Repair	2	2	2
7. Fabric	1	2	2
8. Assemble	1	1	2
9. Enormous	0	2	2
10. Conceal	0	2	2
11. Sentence	0	0	2
12. Consume	0	1	2
13. Regulate	0	0	2
14. Terminate		0	2
15. Commerce		0	2
16. Domestic		0	2
17. Tranquil		0	0
18. Ponder			1
19. Designate			1
20. Relucant			1
21. Obstruct			0
22. Sanctuary			0
23. Compassion			0
24. Evasive			0
25. Remorse			0
•			
•			
•			
•			
35. Tirade			
Total	11	20	35

Note.   2 = *understand word*, 1 = *partial understanding*, 0 = *no understand-
ing.* This test is discontinued after five "*no understanding*" or "0" responses.
These response patterns are contrived illustrations based on average subscale ex-
pectations derived from the WAIS Manual (Wechsler, 1981).

(1990–1991) Sidney Hook Memorial Award Address, subtitled, "On the self-
suppression of academic freedom." Here, Coleman outlined some neglected as-
pects in his own research, which likely served to maintain (iatrogenic) educational
policies contributing to the underachievement of both Black and White students.
His treatment expands to include what he considers one of the largest threats to
academic freedom: the threat posed by faculty members themselves, through the
norms they establish about what kinds of questions may be raised in research and
what kinds of questions may not be raised. Examples of such norms are revealed

by the disdain engendered when someone considers biological antecedents to race or sex differences in ability or vocational preferences, and by the often equally intense disdain directed today toward anyone who might consider nonbiological social influences relevant to the development of sexual preference. Coleman maintains that these threats, emanating from faculty members, motivate researchers to neglect inconvenient questions and that they far outweigh threats posed by the usual villains of academic freedom. Coleman's analysis is supported by Erich Fromm:

> The intellectual has one prime task to fulfill, first, last, and always. It is his job to search out the truth as best he can and to speak that truth. It is not the intellectual's primary calling, it is not his primary function, to draft political platforms . . . But it is the intellectual's special task—and this is what defines his role or should define it—to pursue the truth without compromise and without regard for his own or anyone else's interests. If intellectuals restrict their function of finding and speaking the whole truth in the service of any program or any political goals, no matter how praiseworthy the program or the goals may be, then those intellectuals are failing in their own unique task and, ultimately, in the most important *political* task they have. For I feel that political progress depends on how much of the truth we know, how clearly and boldly we speak it, and how great an impression it makes on other people. (Fromm, 1956, p. 116)

It is time that measures of general intelligence be given the opportunity to reveal fully the scope of their scientific and practical significance. They should be incorporated into broad-spectrum epidemiological and social science investigations of human phenomena.[8] The sources cited throughout this special issue collectively converge on the conclusion that the construct of general intelligence captures the most scientifically significant source of human variation uncovered by psychological science. General intelligence is well established as one of the most significant phenotypic human attributes. It is the behavioral attribute that distinguishes us most from other species. Human intelligence accounts for our ability to transform our environment, to create culture and to record as well as transmit culture to future generations. For general intelligence to remain unassimilated into much of the social sciences is scientifically indefensible. For many disciplines, it may even be scientific malpractice.

---

[8]Yet, the scientific import of measures of general intelligence is destined to be underappreciated if those measures are correlated with social phenomena within highly restricted ranges of ability, where people have already been segregated based on general intelligence. This practice is reminiscent of Brand's (1987) report of an investigator who "didn't believe in individual differences," but who rarely studied anyone outside the narrow 110–130 IQ range. Epidemiological and social science inquiry must precisely report and correct for restriction in range of intellectual talent.

REFERENCES

Achter, J.A., Lubinski, D., & Benbow, C.P. (1996). Multipotentiality among intellectually gifted: "It was never there and already it's vanishing." *Journal of Counseling Psychology, 43,* 65–76.

Adler, N.E., Boyce, T., Chesney, M.A., Cohen, S., Folkman, S., Kahn, R.L., & Syme, S.L. (1994). Socioeconomic status and health: The challenge of the gradient. *American Psychologist, 49,* 15–24.

Adler, N.E., & Matthews, K. (1994). Health psychology: Why do some people get sick and some stay well? *Annual Review of Psychology, 45,* 229–259.

Alberts, B. (1994, 21 November). [Letters]. *Newsweek.*

American Psychological Society (1996, April). Human Capital Initiative: Doing the right thing: A research plan for healthy living. *Observer,* Report 4 [Special issue]. Washington, DC: American Psychological Society.

Anastasi, A. (1958). *Differential psychology: Individual and group differences in behavior.* New York: Macmillan.

Anderson, N.B., & Armstead, C.A. (1995). Toward understanding the association of socioeconomic status and health: A new challenge for the biopsychosocial approach. *Psychosomatic Medicine, 57,* 213–225.

Becker, W.C., & Carnine, D.W. (1980). Direct instruction: An effective approach to educational intervention with the disadvantaged and low performers. In B.B. Lahey & A.K. Kazdin (Eds.), *Advances in clinical and child psychology* (Vol. 3). New York: Plenum.

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

Bouchard, T.J., Jr. (1995). Breaking the last taboo [Review of the book *The bell curve: Intelligence and class structure in American life*]. *Contemporary Psychology, 40,* 415–418.

Bouchard, T.J., Jr. (1996). IQ similarity in twins reared apart: Findings and responses to critics. In R.J. Sternberg & E.L. Grigorenko (Eds.), *Intelligence: Heredity and environment.* New York: Cambridge University Press.

Brand, C. (1987). The importance of general intelligence. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy.* New York: Falmer.

Brody, N. (1992). *Intelligence* (2nd ed.). New York: Academic.

Carnap, R. (1950). *Logical foundations of probability.* Chicago: University of Chicago Press.

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

Castell, A. (1935). *A college logic.* New York: Macmillan.

Cleary, T.A., Humphreys, L.G., Kendrick, S.A., & Wesman, A. (1975). Educational uses of tests with disadvantaged students. *American Psychologist, 30,* 15–41.

Cohen, C. (1996). Race, lies, and "Hopwood." *Commentary, 6,* 39–45.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Coleman, J.S. (1990–1991). The Sidney Hook Memorial Award Address: On the self-suppression of academic freedom. *Academic Questions, 4,* 17–22.

Coleman, J.S., Campbell, E.O., Hobson, C.J., McPartland, J., Moody, A.M., Weinfield, F.D., & York, R.L. (1966). *Equality of educational opportunity.* Washington, DC: U.S. Government Printing Office.

Cronbach, L.J. (1957). Two disciplines of scientific psychology. *American Psychologist, 12,* 671–684.

Cronbach, L.J., & Snow, R.E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions.* New York: Irvington.

Davis, B.D. (1976). *Academic standards in medical schools. The New England Journal of Medicine, 294,* 118–119.

Davis, B.D. (1986). *Storm over biology: Essays on science, sentiment, and public policy.* Buffalo: Prometheus.

Dawis, R.V. (1991). Vocational interests, values, and preferences. In M.D. Dunnette & L.M. Hough (Eds.), *Handbook of industrial and organizational psychology,* (Vol. 2). Palo Alto: Consulting Psychologist Press.

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

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.

Englemann, S., & Carnine, D. (1982). *Theory of instruction: Principles and applications.* New York: Irvington.

Eysenck, H.J. (1970). *Crime and personality* (2nd ed.). London: Granada.

Eysenck, H.J. (1979). *The structure and measurement of intelligence.* New York: Springer-Verlag.

Eysenck, H.J. (1995). *Genius: The natural history of creativity.* Cambridge: Cambridge University Press.

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

Flynn, J.R. (1984). The mean IQ of Americans: Massive gains 1932–1978. *Psychological Bulletin, 95,* 29–51.

Flynn, J.R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101,* 171–191.

Friedman, G.D. (1994). *Primer of epidemiology* (4th ed.). New York: McGraw-Hill.

Fromm, E. (1956). *The art of loving.* New York: Bantam.

Gordon, R.A. (1987). SES versus IQ in the race–IQ–delinquency model. *International Journal of Sociology and Social Policy, 7,* 30–96.

Gordon, R.A. (1988). Thunder from the left. *Academic Questions, 1*(3), 74–92.

Gordon, R.A. (1997). Everyday life as an intelligence test: Effects of intelligence and intelligence context. *Intelligence, 24*(1), 203–320.

Gottfredson, L.S. (Ed.). (1986). The *g* factor in employment [Special issue]. *Journal of Vocational Behavior, 29*(3).

Gottfredson, L.S. (1997a). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence, 24*(1), 13–23.

Gottfredson, L.S. (1997b). Why *g* matters: The complexity of everyday life. *Intelligence, 24*(1), 79–132.

Gottfredson, L.S., & Sharf, J.C. (Eds.). (1988). Fairness in employment testing [Special issue]. *Journal of Vocational Behavior, 33*(3).

Gould, S.J. (1994, November 28). Curveball [Review of the book *The Bell Curve*]. *New Yorker,* pp. 139–149.

Green, B.F. (1978). In defense of measurement. *American Psychologist, 33,* 664–670.

Hedges, L.V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high scoring individuals. *Science, 269,* 41–45.

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

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,* 105–120.

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). Limited vision in the social sciences. *American Journal of Psychology, 104,* 333–353.

Humphreys, L.G. (1994). Intelligence from the standpoint of a (pragmatic) behaviorist. *Psychological Inquiry, 5,* 179–192.

Humphreys, L.G., Davie, T.C., & Kashima, E. (1986). Experimental measures of cognitive privilege/deprivation and some of their correlates. *Intelligence, 10,* 355–376.

Humphreys, L.G., & Lubinski, D. (1996). Brief history and psychological significance of spatial-visualization abilities. In C.P. Benbow and D. Lubinski (Eds.), *Intellectual talent: Psychometric and social issues.* Baltimore: Johns Hopkins University Press.

Humphreys, L.G., Lubinski, D., & Yao, G. (1993a). 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., Lubinski, D., & Yao, G. (1993b). Correlates of some curious regressions on a measure of general intelligence. *Journal of School Psychology, 31,* 383–405.

Humphreys, L.G., & Swets, J.A. (1991). Comparison of predictive validities measured with biserial correlations and ROCs of signal detection theory. *Journal of Applied Psychology, 76,* 316–321.

Hunter, J.E., Schmidt, F.L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin, 86,* 721–735.

Jenkins, J.J., & Paterson, D.G. (1961). *Studies in individual differences: The search for intelligence.* New York: Appleton-Century-Crofts.

Jensen, A.R. (1980). *Bias in mental testing.* New York: Free Press.

Jensen, A.R. (1991). Spearman's g and the problem of educational equity. *Oxford Review of Education, 17,* 169–187.

Jensen, A.R., & Weng, L. (1994). What is a good g? *Intelligence, 18,* 231–258.

Kuttner, R.E., & Lorincz, A.B. (1968). Utilization of accentuated environmental inequalities in research on racial differences. *Science, 160,* 439–440.

Lerner, B. (1991). Good news about American education. *Commentary, 91,* 19–25.

Lerner, B. (1995, March 6). Aim higher. *National Review,* 56–60.

Loehlin, J.C., Lindzey, G., & Spuhler, J.M. (1975). *Race differences in intelligence.* San Francisco: Freeman.

Lubinski, D. (1996). Applied individual differences research and its quantitative methods. *Psychology, Public Policy, and Law, 2,* 187–203.

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 (SMPY): The first three decades of a planned 50-year study of intellectual talent. In R. Subotnik & K. Arnold (Eds.), *Beyond Terman: Longitudinal studies in contemporary gifted education.* Norwood, NJ: Ablex.

Lubinski, D., Benbow, C.P., & Ryan, J. (1995). Stability of vocational interests among the intellectually gifted from adolescence to adulthood: A 15-year longitudinal study. *Journal of Applied Psychology, 80,* 90–94.

Lubinski, D., Benbow, C.P., & Sanders, C.E. (1993). Reconceptualizing gender differences in achievement among the gifted. In K.A. Heller, F.J. Monks, & A.H. Passow (Eds.), *International handbook for research on giftedness and talent.* Oxford: Pergamon.

Lubinski, D., & Dawis, R.V. (1992). Aptitudes, skills, and proficiencies. In M. Dunnette & L.

Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 3). 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.

Lubinski, D., & Humphreys, L.G. (1992). Some bodily and medical correlates of mathematical giftedness and commensurate levels of socioeconomic status. *Intelligence, 16,* 99–115.

Lubinski, D., & Humphreys, L.G. (1996). Seeing the forest from the trees: When predicting the behavior of status of groups, correlate means. *Psychology, Public Policy, and Law, 2,* 363–376.

Lubinski, D., Schmidt, D.B., & Benbow, C.P. (1996). A 20-year stability analysis of the Study of Values for intellectually gifted individuals from adolescence to adulthood. *Journal of Applied Psychology, 81,* 443–451.

Lykken, D.T. (1995). *The antisocial personalities.* Hillsdale, NJ: Erlbaum.

Lynn, R. (1987). The intelligence of the Mongoloids: A psychometric, evolutionary and neurological theory. *Personality and Individual Differences, 8,* 813–844.

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

Meehl, P.E. (1970). Nuisance variables and the ex post facto design. In M. Radner & S. Winokur (Eds.), *Minnesota Studies in the Philosophy of Science* (IV). Minneapolis: University of Minnesota Press.

Meehl, P.E. (1971). High school yearbooks: A reply to Schwarz. *Journal of Abnormal Psychology, 77,* 143–148.

Meehl, P.E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry, 1,* 108–141, 173–180.

Morris, J.N. (1967). *Use of epidemiology.* London: E.S. Livingston.

National Center for Health Statistics. (1987). Characteristics of American Indian and Alaska Native births: United States 1984. *Monthly Vital Statistics Report, 36,* No. 3. Hyattsville, MD: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Health Statistics.

National Center for Health Statistics. (1994a). Advance report of final natality statistics, 1992. *Monthly Vital Statistics Report, 43,* No. 5, Supplement. Hyattsville, MD: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Health Statistics.

National Center for Health Statistics. (1994b). Advance report of final natality statistics, 1992. *Monthly Vital Statistics Report, 43,* No. 6, Supplement. Hyattsville, MD: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Health Statistics.

Nichols, D.S., & Marks, P.A. (1992). *Reflections and remembrances: A tribute to Starke R. Hathaway.* Minneapolis: University of Minnesota Press.

Plomin, R. (1994). *The nature of nurture: Family environment.* Beverly Hills: Sage.

Plomin, R., & Petrill, S.A. (1997). Genetics and intelligence: What's new? *Intelligence, 24*(1), 53–77.

Rest, J.R., & Narvaez, D. (1994). *Moral development in the professions: Psychology and applied ethics.* Hillsdale, NJ: Erlbaum.

Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist, 25,* 775–777.

Rowe, D.C. (1994). *The limits of family influence.* New York: Guilford.

Roznowski, M. (1987). Use of tests manifesting sex differences as measures of intelligence: Implications for measurement bias. *Journal of Applied Psychology, 72,* 480–483.

Sanders, C.E., Lubinski, D., & Benbow, C.P. (1995). Does the Defining Issues Test measure psychological phenomena distinct from verbal ability?: An examination of Lykken's query. *Journal of Personality and Social Psychology, 69,* 498–504.

Sarich, V. (1990). The institutionalization of racism at Berkeley. *Academic Questions, 4,* 72–81.

Scarr, S. (1992). Developmental theories for the 1990's: Development and individual differences. *Child Development, 63,* 1–19.

Schmidt, F.L. (1988). The problem of group differences in ability test scores in employment selection. *Journal of Vocational Behavior, 33,* 272–292.

Seligman, D. (1992). *A question of intelligence: The IQ debate in America.* New York: Birch Lane Press.

Snow, R.E. (1989). Aptitude-treatment interaction as a framework for research on individual differences in learning. In P.L. Ackerman, R.J. Sternberg, & R.G. Glasser (Eds.), *Learning and individual differences: Advances in theory and research.* New York: Freedman.

Snyderman, M., & Rothman, S. (1987). Survey of expert opinion on intelligence and aptitude testing. *American Psychologist, 42,* 137–144.

Snyderman, M., & Rothman, S. (1988). *The IQ controversy, the media and public policy.* New Brunswick, NJ: Transaction Books.

Sowell, T. (1992). *Inside American education: The decline, the deception, the dogmas.* New York: Free Press.

Spearman, C. (1927). *Abilities of man: Their nature and measurement.* New York: Macmillan.

Stanley, J.C. (1971). Predicting college success of the educationally disadvantaged. *Science, 171,* 640–647.

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. Ambroson (Eds.), *Talent development.* Unionville, NY: Trillium Press.

Stewart, G.T. (1972). *Trends in epidemiology: Application to health service research and training.* Springfield, IL: Charles C. Thomas.

Taylor, H.C., & Russell, J.T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology, 23,* 565–578.

Taylor, S.E. (1991). *Health psychology* (2nd ed.). New York: McGraw-Hill.

Terman, L.M., & Merrill, M.A. (1960). *Stanford-Binet Intelligence Scale: Manual for the third revision, Form L-M.* Boston: Houghton Mifflin.

Thorndike, E.L. (1911). *Individuality.* Cambridge, MA: Riverside Press.

Tyler, L.E. (1965). *The psychology of human differences* (3rd ed.). New York: Appleton-Century-Crofts.

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

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

Vernon, P.E. (1979). *Intelligence: Heredity and environment.* San Francisco: Freedman.

Vernon, P.E. (1982). *The abilities and achievements of Orientals in North America.* New York: Academic Press.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children.* New York: Psychological Corporation.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale-Revised.* New York: The Psychological Corporation.

Wigdor, A.K., & Garner, W.R. (Eds.) (1982). *Ability testing: Uses, consequences, and controversies.* Washington, DC: National Academy Press.

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

Williams, M.V., Parker, R.M., Baker, D.W., Parikh, N.S., Pitkin, K., Coates, W.C., & Nurss, J.R. (1995). Inadequate functional health literacy among patients at two public hospitals. *Journal of the American Medical Association, 274,* 1677–1682.

Williamson, E.G. (1965). *Vocational counseling: Some historical, philosophical, and theoretical perspectives.* New York: McGraw-Hill.

Wilson, J.Q., & Herrnstein, R.J. (1985). *Crime and human nature.* New York: Simon & Schuster.

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