
Survey of Expert Opinion on Intelligence and Aptitude Testing

Mark Snyderman
Stanley Rothman

Harvard University
Smith College

ABSTRACT: *Psychologists and educational specialists with expertise in areas related to intelligence testing responded to a questionnaire dealing with a wide variety of issues constituting the IQ controversy. Overall, experts hold positive attitudes about the validity and usefulness of intelligence and aptitude tests. Tests are seen as adequately measuring most important elements of intelligence, although the tests are believed to be somewhat racially and socioeconomically biased. There is overwhelming support for a significant within-group heritability for IQ, and a majority of respondents feel that black-white and socioeconomic status IQ differences are also partially hereditary. Problems with intelligence tests are perceived in the influence of nonintellectual characteristics on test performance and in the frequent misinterpretation and overreliance on test scores in elementary and secondary schools. Despite these difficulties, experts favor the continued use of intelligence and aptitude tests at their present level. Variation in responding to substantive questions on testing is largely resistant to prediction by a host of demographic and background variables, including within-sample variation in expertise.*

Intelligence tests have been under attack practically since their inception (Cronbach, 1975; Haney, 1981). Critics have claimed, among other things, that intelligence and aptitude tests measure nothing but test-taking skills, have little predictive power, are biased against certain racial and economic groups, are used to stigmatize low scorers, and are tools developed and fostered by those in power in order to maintain the status quo (see Block & Dworkin, 1976, and Houts, 1977, for collections of such critiques). Though perhaps not as apparent as 10 (or 60) years ago, such criticisms remain prevalent (e.g., Gould, 1981; Lewontin, Rose, & Kamin, 1984; Owen, 1985). Moreover, critics of testing appear to have much influence in such organizations as the National Education Association, the news media, the New York State Legislature, and the courts (Bersoff, 1981; Herrnstein, 1982; Lerner, 1980).

It is not surprising, of course, in light of the important role intelligence and aptitude tests play in the allocation of valued resources and opportunities, that testing has been a topic of concern in the popular press and in all three branches of government. What is surprising is that much of the public controversy seems to be uninformed. Those who must reach policy decisions about

testing often seem more influenced by political considerations than by the empirical literature.

There is, of course, no shortage of appeal to expertise. Public opinion and policy are influenced by the perception of expert opinion: Witness the standard procedure in Congressional hearings and news media stories on technical issues. In public forums, the impression is often given by those who attack tests (e.g., CBS News, 1975; *Larry P. v. Wilson Riles*, 1979) that many of the long-accepted "facts" about intelligence tests are subjects of great dispute within the expert community or that most experts actually agree, for example, that tests are culturally biased, meaningless as anything but predictors of success in school, and unrelated to an individual's genetic endowment. These claims may very well be true, but they are rarely made with sufficient supporting evidence. It is important, therefore, to try to assess the veracity of assertions that there is substantial controversy about, and even animosity toward, testing among those most familiar with the empirical evidence.

Surveys of opinion on intelligence testing and related issues, among any group, have been scarce since the advent of the most recent wave (post 1969) of testing criticism (see Brim, Glass, Neulinger, & Firestone, 1969, for an earlier comprehensive survey of public opinion, and Lerner, 1981, for a review of more recent public opinion surveys). One group that has been particularly ill served by survey research is testing experts. Those who conduct research on the nature of intelligence and test use and those who design and validate tests, and who therefore are most qualified to evaluate criticisms of testing in the context of the body of psychometric and cognitive ability literature have rarely been asked their opinions about the most important issues of public contention surrounding intelligence tests. To date, there are no comprehensive polls of this sort.

Such a survey is needed, but not because it will resolve any of the various controversies surrounding testing; issues of fact are not settled via consensus. A comprehensive survey of expert opinion about intelligence testing is necessary because the use of intelligence and aptitude testing represents an important public policy issue. A survey of expert opinion will not settle this issue; but it will allow a clearer picture of informed opinion to enter the public debate. In a way, it is a method of pooling "expert testimony" for the benefit of those charged with policy decisions. It should also allow anyone interested

in the IQ controversy to achieve a better understanding of the issues involved.

Method

Subjects

The composition of the survey sample is described in Table 1. The purpose of this research was to survey expert opinion about the IQ controversy. Because the controversy is a broad one, the population that constitutes "experts" is not immediately apparent. It was therefore necessary to define the population through the various considerations that guided sample selection. There were three primary considerations. First, the population should be neither so broad as to contain a large proportion of individuals with little or no experience with intelligence testing nor so narrow as to include only those who might be considered to have a vested interest in testing. Second, we wished to include individuals with a variety of perspectives on the problem, including those who might have expertise on only a small part of the controversy. For this purpose, we divided the population into primary and secondary groups. Primary groups were those professional organizations whose members might be expected to be knowledgeable on a variety of IQ-related topics. Secondary groups were organizations whose members were likely to know testing from only a narrow perspective. For example, members of the American Sociological Association (ASA) who identify themselves as sociologists of education, were included for their expertise on the role of testing in society, and members of the Cognitive Science Society, were included for their expertise on the nature of intelligence and cognitive abilities.

The final criterion was that the population, and the sample drawn therefrom, be weighted in favor of those with the most expertise, as indicated by research and publications on issues dealing with testing. Therefore, only scholarly organizations were sampled. The sample was also weighted toward those organizations, and those members within the organizations, thought to have the most expertise. Because members of primary groups were believed to have more overall expertise than members of secondary groups, twice as many members were selected from each primary group as from each secondary group. For those organizations where it was possible to separate PhD from non-PhD members, only members with doctorates were sampled. Within each division of the APA, despite the fact that there are far fewer Fellows than Members, half of the sample was drawn from Fellows, and half from Members.

The sample was drawn randomly from the most recent available membership directory of each of the organizations. Many of those in the sample are, of course, members of more than one of the listed organizations or American Psychological Association (APA) divisions, but the sample was chosen so that there was no overlap between groups. The final sample consisted of 1,020 social scientists and educators.

Table 1
Composition of Survey Sample

Groups	N
Primary Groups	
American Educational Research Association	120
National Council on Measurement in Education	120
American Psychological Association:	
Developmental Psychology	60 Fellows 60 Members
Educational Psychology	60 Fellows 60 Members
Evaluation and Measurement	60 Fellows 60 Members
School Psychology	60 Fellows 60 Members
Secondary groups	
American Sociological Association:	
Education	60
Behavior Genetics Association	60
Cognitive Science Society	60
American Psychological Association:	
Counseling Psychology	30 Fellows 30 Members
Industrial/Organizational Psychology	30 Fellows 30 Members
Total	1,020

Materials

The questionnaire was an 8½ × 11 in. 16-page booklet containing 48 questions, many with multiple parts, divided into six sections. (The questions discussed in the Discussion section represent only the key findings.)¹ Four of the sections contained substantive questions about

This research is part of a larger study on Politics and Science, directed by Stanley Rothman and sponsored by the Center for the Study of Social and Political Change at Smith College and the Research Institute on International Change, Columbia University.

We thank James Davis, Richard Herrnstein, Sheldon White, and the pilot study respondents for their helpful comments during the development of the questionnaire; Kathleen McCartney and Douwe Yntema for their help with the statistical analyses; and Jacqueline Boudreau and Arianne Stubbs for their assistance in the preparation and mailing of the questionnaires and follow-ups.

Correspondence concerning this article should be addressed to Mark Snyderman, 5110 S. Kenwood Ave., Apt. No 508, Chicago, IL 60615.

¹ Questions not described in the Discussion section fall into one of the following three categories: (a) those related more to the public and political aspects of the IQ controversy than to testing itself, for example, ratings of news media accuracy on coverage of testing issues and ratings of opinion about key scientists involved in the controversy; (b) questions of secondary importance, such as validity ratings of particular admissions tests and the use of IQ tests for placement of educable mentally retarded students; and (c) questions whose answers (i.e., a large number of written responses indicating confusion) demonstrate sufficient ambiguity to prevent meaningful interpretation. Complete results from all questions, as well as codebooks and computer tapes, will be deposited at the Roper Public Opinion Research Center at the University of Connecticut Storrs, Connecticut, for examination by interested scholars.

intelligence and testing, and two asked about various demographic and background characteristics of the respondents. The scope of the substantive questions was intended to include most areas of contention within the relevant academic literature, with an emphasis on areas of particular concern in the public debate.

Procedure

In September of 1984, following pretesting, 1,020 questionnaires were mailed. Each envelope contained a questionnaire, a stamped return envelope, and a cover letter. The cover letter contained an explanation of the purpose of the questionnaire (to help clarify confusion over testing), its importance in light of the widespread use and controversy over tests, and an assurance of complete confidentiality (the questionnaire itself contained an identification number for the purposes of follow-up mailings). Because many respondents were not expected to have expertise in all areas of testing, the cover letter asked subjects to check the NQ ("not qualified") response for any question they did not feel qualified to answer. This category also served for "no response/don't know."

Approximately two weeks after the initial mailing, postcard reminders were sent to all the subjects who had not yet responded. About four weeks later, a second set of questionnaires were sent out to the remaining nonrespondents. The final response tally contained 661 completed questionnaires (65%). Forty-nine subjects returned their questionnaires, indicating they were not qualified to answer any of the substantive questions. Seventeen subjects were deceased or were otherwise incapacitated, and 27 subjects simply returned their questionnaires unanswered with no explanation. There was little variation in response rate between groups within the sample.

Two hundred sixty-six (26%) of the questionnaires were not returned at all. Phone calls were made to 40 (15%) of these nonrespondents in order to determine if they differed in any important way from respondents and to find out their reasons for nonresponse. These subjects were asked some of the more important substantive and demographic questions, but response rates were often quite low; these were individuals who already had not responded to three mailings. Their responses to questions for which there were a sufficient number of answers for meaningful comparison (at least 50% response rate) were not significantly different from those of respondents to the mailed questionnaire. More informative perhaps were the reasons these subjects gave for not responding. All 40 of these subjects answered this question. Twenty-three said that they were too busy to respond, and 12 did not feel qualified. Only 6 expressed any aversion to the questionnaire itself (respondents could give more than one reason). In all, given the nature of responses received from the phone-call sample, and their reasons for not responding to the mailed questionnaire, there seems little reason to believe that the results would look significantly different had the entire sample of 1,020 participated.

Discussion

Professional Activities and Involvement With Intelligence Testing

The degree of expertise about intelligence and testing varies widely among respondents, but, on the whole, the sample is adequately characterized as expert. Approximately half of all respondents are faculty members at a college or university, and the bulk of the remainder classify themselves as psychologists or educational specialists working in some other capacity. Fifty-five percent are planning or carrying out research in some area related to intelligence or intelligence testing. The most common areas of research are the nature of intelligence, test development and validation, and testing in elementary and secondary schools.

Sixty-seven percent of respondents have written at least one article or chapter related to intelligence or testing, and 57% have given at least one such speech or lecture to other than a classroom audience during the past two years. The mean number of articles written is 11 ($Mdn = 3$), with articles written for an academic/professional audience about five times more common than those written for a general audience. The most common article topics parallel those for areas of research.

The Nature of Intelligence

1. *Consensus.* Respondents were asked whether they agreed that there is a consensus among psychologists and educators as to the kinds of behaviors that are labeled "intelligent." The argument represents Cleary, Humphreys, Kendrick, and Wesman's (1975) response to the criticism that "intelligence" is not well-defined. A majority of respondents (though, ironically, not a consensus) agree that there is a consensus. Fifty-three percent either somewhat or strongly agree, compared to 39.5% who either somewhat or strongly disagree. The remaining 7.5% do not respond to the question.

2. *Important elements of intelligence.* This question constituted a more direct attempt to determine if a consensus exists, at least at the conceptual level. Respondents were asked to check all behavioral descriptors listed (there were 13, and space for writing in others) that they believe to be an important element of intelligence. Results are shown in the first data column of Table 2. Response rate (r.r.) is 93%. Descriptors fall into one of three well-defined categories: those for which there is near unanimity (> 96% agreement among those who answered the question)—"abstract thinking or reasoning," "the capacity to acquire knowledge," and "problem solving ability"; those checked by a majority of respondents (60%–80%)—"adaptation to one's environment," "creativity," "general knowledge," "linguistic competence," "mathematical competence," "memory," and "mental speed"; and those rarely checked (< 25%)—"achievement motivation," "goal-directedness," and "sensory acuity." No descriptors were added to the list by more than 2% of respondents.

3. *Important elements not measured.* Respondents were asked to check each of the behavioral descriptors

that they believe to be an important element of intelligence but that they do not feel is adequately measured by the most commonly used intelligence tests. These results are also given in Table 2. Response rate is 87%. This question essentially concerns construct validity, and there appears to be substantial support among experts for the validity of the most commonly used intelligence tests. Of the 10 behavioral descriptors checked as important elements by more than 60% of respondents, only 2, "adaptation to one's environment" and "creativity," were checked by a majority as not adequately measured, and only 1 other, "capacity to acquire knowledge," was checked by more than 28%. The problems with "adaptation to one's environment" reflect the common criticism that tests are much better at measuring traits important to success in school than general life skills. Similarly, the "creativity" result is consistent with the poor correlation between tests of intelligence and tests of creativity (Sattler, 1982). Somewhat troublesome for supporters of testing is the fact that 42% of those who believe "capacity to acquire knowledge" is an important element of intelligence, which includes virtually all respondents, do not believe it is adequately measured by intelligence tests.

4. *The importance of personal characteristics to intelligence test performance.* Respondents were asked to rate each of six personal characteristics for their importance to performance on intelligence tests. Ratings were made on a 4-point scale, where 1 was *of little importance*, and 4 was *very important*. All of these essentially non-intellectual characteristics are seen as at least somewhat important to test performance. Mean ratings are as fol-

lows: achievement motivation, 2.87 ($SD = 0.964$, $r.r. = 91.5\%$); anxiety, 2.68 ($SD = 0.901$, $r.r. = 90.6\%$); attentiveness, 3.39 ($SD = 0.744$, $r.r. = 92.6\%$); emotional lability, 2.52 ($SD = 0.938$, $r.r. = 83.2\%$); persistence, 2.96 ($SD = 0.872$, $r.r. = 91.2\%$); and physical health, 2.34 ($SD = 0.892$, $r.r. = 92\%$).

5. *General intelligence.* This question asked, "Is intelligence, as measured by intelligence tests, better described in terms of a primary general intelligence factor and subsidiary group and special ability factors, or entirely in terms of separate faculties?" Despite the so-called "arbitrariness" of factor analytic solutions, most respondents are able to reach a decision on how most meaningfully to describe intelligence test results. Fifty-eight percent favor some form of a general intelligence solution, whereas 13% feel separate faculties are superior. Only 16% think the data are sufficiently ambiguous as to not favor either solution.

The Heritability of IQ

6. *Sources of heritability evidence.* The claim has been made, most notably by Kamin (1974), that there is no reasonable evidence for a nonzero heritability of IQ. Respondents were presented with a list of five sources of evidence and asked to check all sources that they believe provide reasonable support for a significant nonzero heritability of IQ in the American white population. Sources included kinship correlations, studies of monozygotic (MZ) twins reared apart, monozygotic-dizygotic twin comparisons, twin family studies, and adoption studies. Twenty-five percent of subjects did not feel qualified to answer this question. Of those who did respond, 94% checked at least one source of evidence, and none of the sources was checked by less than half of the respondents. Support is greatest for studies of MZ twins reared apart (84.4%) and weakest for twin family studies (55.3%). The latter result is understandable because twin family studies are a relatively recent development in the behavior genetics of IQ (Scarr & Carter-Saltzman, 1982). Taken together, these results are a strong indication that experts believe within-group differences in IQ to be at least partially inherited.

7. *White heritability estimate.* Despite a consensus that there is a significant heritability to IQ in the American white population, experts disagree on the issue of whether there is sufficient evidence to arrive at a reasonable estimate of this heritability. Thirty-nine percent feel that there is sufficient evidence, compared to 40% who do not. Twenty-one percent do not feel qualified to answer. Only those respondents who feel there is sufficient evidence were asked to provide a heritability estimate. The mean estimate for the 214 received is 0.596 ($SD = 0.166$), meaning that these experts believe, on the average, that 60% of the variation in IQ within the American white population is associated with genetic variation.

8. *Black heritability estimate.* Experts are much less inclined to believe that sufficient evidence exists for an estimate of IQ heritability among the American black population. Twenty percent feel there is sufficient evi-

Table 2
Important Elements of Intelligence

Descriptor	% of respondents checking as important	% of respondents checking as not adequately measured*
Abstract thinking or reasoning	99.3	19.9
Problem-solving ability	97.7	27.3
Capacity to acquire knowledge	96	42.2
Memory	80.5	12.7
Adaptation to one's environment	77.2	75.3
Mental speed	71.7	12.8
Linguistic competence	71	14
Mathematical competence	67.9	12.1
General knowledge	62.4	10.7
Creativity	59.6	88.3
Sensory acuity	24.4	57.7
Goal-directedness	24	64.1
Achievement motivation	18.9	71.7

* Respondents include only those who had previously indicated that the descriptor is an important element of intelligence.

dence, and 54% feel there is not. The mean heritability estimate for 101 received is 0.571 ($SD = 0.178$).

The large percentage of respondents who indicated that they do not feel qualified to answer questions on the heritability of IQ is testimony to the highly technical nature of the topic. Despite the self-selection of experts, a further comparison was made between members of the Behavior Genetics Association (BGA; $N = 34$) and the rest of the sample on the heritability questions. The only significant difference between these groups is on the question of sufficient evidence for a white heritability estimate. BGA members are much more likely to believe that sufficient evidence exists than are nonmembers (76% vs. 37%), $\chi^2(1, N = 34) = 10.41, p < .002$, two-tailed. There is no difference in the heritability estimates given, however.

Race, Class, and Cultural Differences in IQ

9. *Racial bias.* All of the questions on test bias asked for a rating on a 4-point scale, where 1 was described as *not at all or insignificantly biased*, 2 was *somewhat biased*, 3 was *moderately biased*, and 4 was *extremely biased*. This question asked to what extent the most commonly used intelligence tests are biased against American blacks. Bias was defined as an average black American's test score underrepresenting his or her actual level of those abilities the test purports to measure, relative to the average ability level of members of other racial and ethnic groups. The mean bias rating for this question is 2.12 ($SD = 0.787$, $r.r. = 84.1\%$), indicating that experts believe there to be some racial bias in intelligence tests, but less than what would be considered a moderate amount.

10. *Economic bias.* This question is identical to that on general racial bias, except it asks about bias against lower socio-economic groups rather than against blacks. The mean bias rating is slightly higher than for racial bias, at 2.24 ($SD = 0.813$, $r.r. = 84.7\%$).

11. *Other biasing factors.* Respondents were presented with a list of five factors that have been proposed at various times as differentially affecting the test scores of members of certain ethnic, racial, or economic groups. Mean bias ratings are as follows: race of the examiner, 1.91 ($SD = 0.758$, $r.r. = 85.9\%$); language and dialect of the examiner, 2.46 ($SD = 0.865$, $r.r. = 86.2\%$); attitude of the examiner toward the group in question, 2.74 ($SD = 0.932$, $r.r. = 85.6\%$); test taker anxiety, 2.63 ($SD = 0.894$, $r.r. = 85.1\%$); and test taker motivation, 2.91 ($SD = 0.925$, $r.r. = 85.6\%$). The substantial ratings for many of these items parallel the belief in the influence of nonintellectual personal characteristics noted in question 4.

12. *The source of the black-white difference in IQ.* This is perhaps the central question in the IQ controversy. Respondents were asked to express their opinion of the role of genetic differences in the black-white IQ differential. Forty-five percent believe the difference to be a product of both genetic and environmental variation, compared to only 15% who feel the difference is entirely due to environmental variation. Twenty-four percent of experts do not believe there are sufficient data to support any reasonable opinion, and 14% did not respond to the

question. Eight experts (1%) indicate a belief in an entirely genetic determination.

13. *The source of socioeconomic class differences in IQ.* The case for genetic determination is even more strongly felt for socioeconomic status (SES) differences. Fifty-five percent of experts choose the genetic-environmental option, as opposed to 12% for strictly environmental. Eighteen percent do not feel there are sufficient data, and 15% were nonrespondents. Only one respondent attributes the difference entirely to genetics. This question, as the next one, is relevant to Herrnstein's (1971) thesis that in a society where the abilities measured by intelligence tests are important to success, socioeconomic class differences, particularly differences related to those abilities, will be partially genetic.

14. *Social mobility.* This question asked, "In your opinion, to what degree is the average American's socioeconomic status determined by his or her IQ?" Respondents are generally supportive of the idea of the United States as somewhat of an intellectual meritocracy. Sixty percent feel that IQ is an important, but not the most important, determinant of SES. Twenty-one percent believe IQ plays only a small role in determining SES, and 3% feel it is not at all important. Only 2% rate IQ as the most important determinant of SES, and 14% were nonrespondents.

The Use of Intelligence Testing

15. *Frequency of test misuse.* It is not uncommon for those who are otherwise supporters of standardized testing to complain about misuse and misinterpretation of test scores (e.g., Jensen, 1980). This question assessed expert opinion of the prevalence of errors in test use in elementary and secondary schools. Table 3 presents the mean prevalence ratings for each of five types of test misuse. Ratings were made on a 4-point scale, where 1 was *rarely present*, 2 was *sometimes present*, 3 was *often present*, and 4 was *almost always present*. Respondents believe all types of misuse to be at least sometimes present, with the highest ratings received for instances of overuse or overreliance on test scores that stem from ignoring test inaccuracies.

16. *Test use.* For each of seven common intelligence and aptitude test uses, respondents were asked to indicate the importance they feel such tests should have, relative to the role they now have. Ratings were made on a 7-point scale, where 1 represented a *severely reduced role*, 4 was *remain about the same*, and 7 was *severely increased role*. Mean ratings for each test use are presented in Table 4. With the exception of testing in employment, and to a lesser extent in tracking decisions in elementary and secondary schools, experts seem generally satisfied with the status quo in test use. There appears to be a general belief in the validity of intelligence and aptitude tests for various educational purposes despite the perception that these tests are often misused in elementary and secondary schools.

Those who are conducting research or who have written about employment tests have better things to say

Table 3
Intelligence Test Misuse in Elementary and Secondary Schools

Source	Mean prevalence rating ^a		% responding
	M	SD	
Administration under improper conditions, such as failure to follow prescribed time limits, or in an environment with significant distractors	2.2	.664	76.9
Use of English-language test results for long-range predictions concerning students for whom English is a second language	2.41	.74	71.4
Comparison of test scores between students while ignoring limitations set by test reliability and measurement error	2.8	.76	80.3
Comparison of intelligence and achievement test scores as a measure of under- or overachievement while ignoring test reliability and measurement error, and differences in test domain	2.88	.736	79.3
Use of tests in making decisions for which they have limited or unknown validity	2.75	.747	80.8

^a 1 = Rarely present, 2 = sometimes present, 3 = often present, and 4 = almost always present.

about them than the rest of the expert population. Employment testing experts rate the use of tests for both hiring decisions (4.11 vs. 3.11), $\chi^2(1, N = 121) = 36.67$, $p < .0001$, two-tailed, and promotion decisions (3.56 vs. 2.67), $\chi^2(1, N = 121) = 26.9$, $p < .0001$, two-tailed, significantly higher than do the rest of the sample.

Specific Expertise

One of the primary reasons for sampling from a wide variety of expert groups and for asking about specific topics of research and authorship and other experiences with testing was to examine the effects of more specific expertise on questionnaire responding. For each of Questions 1 through 16, comparisons were made between those whose experiences were of particular relevance and the rest of the sample. Thus, for example, those who were conducting research or who had written on bias in intelligence tests served as specific experts for the test bias questions. Similarly, those involved with admissions tests or research on the nature of intelligence or any of the other topics covered by the questionnaire also served as specific experts. For some of the questions, specific ex-

periences and affiliations, such as having administered a group or individual intelligence test, or being a member of the Cognitive Science Society, also served to classify respondents as experts. Overwhelmingly, the results of these comparisons are not statistically significant. The important exceptions have already been discussed with the general results from each question. Even when these differences are significant, they are not large.

The relative lack of influence of specific expertise may be partially the result of self-selection on the part of respondents. Subjects were asked to respond NQ to all questions that they did not feel qualified to answer. To the degree that subjects were honest in their self-assessments, respondents were even more expert than the sample as a whole. Such restriction of range due to self-selection makes any attempt to account for within-sample variation more difficult.

Principal Component Analysis

To facilitate further analyses, supervariables were created from substantive-question responses via principal component analysis. Four interpretable factors emerged from this analysis, accounting for 12.1%, 11.3%, 9.2%, and 6.3% of the variance. They were labeled Test Usefulness, Test Bias, Personal Characteristics, and Test Misuse. The first factor reveals the following pattern: belief in a consensus about intelligence, belief in the importance of IQ in determining SES, and particularly high loadings for all test uses. The substantial loadings for Factor 2 are almost entirely for the various test bias questions. Factor 3 has high loadings for all of the nonintellectual characteristics in Question 4, as well as for the sections of Question 11 dealing with bias caused by anxiety and motivation. The fourth factor picks up all four sources of test misuse (Question 15) that were included in the analysis. The only question that does not load on any of the four factors is Question 6 on the sources of heritability evidence. This is probably the result of too little variation in responding.

Table 4
Preferred Level of Intelligence and Aptitude Test Use

Use	Mean rating ^a		% responding
	M	SD	
Diagnosis and special education planning in elementary and secondary schools	3.98	1.22	79.6
Tracking decisions in elementary and secondary schools	3.43	1.43	77.9
College admissions	3.94	1.14	85.3
Graduate and professional school admissions	3.96	1.27	84.4
Vocational counseling	4.01	1.32	77.3
Hiring decisions	3.36	1.5	74.3
Promotion decisions	2.89	1.54	73.2

^a 1 = Severely reduced role, 4 = remain about the same, and 7 = severely increased role.

Supervariables were formed corresponding to each of the four factors. Normalized variables were combined using a weighting system such that only variables loading with an absolute value greater than 0.3 on a given factor were combined to form the corresponding supervariable, positive loading variables being added, and negative loading variables subtracted. Questions with loadings of absolute value greater than 0.6 were given double weight. Missing values were coded as zero and included in the supervariables.

The Effects of Demographic and Background Variables

Table 5 presents correlations between (a) various demographic and background variables and (b) the four supervariables. Seventy-two percent of respondents are male, and the mean age of respondents is 52 years. Authorship is intended as a measure of general expertise and is defined as the number of articles or chapters written on testing and related issues. The data in Table 5 indicate that authorship, like age and masculinity, is marginally associated with traditional pro-testing views.

Political perspective represents a composite of two sets of measures. The first is agreement or disagreement with a series of six political statements discovered, in a previous investigation incorporating many more such statements, to load highly on a factor representing overall political perspective (Rothman & Lichter, 1984). The statements dealt with such issues as affirmative action and the desirability of socialism. The second measure was a self-assessment of global political perspective on a 7-point scale, where 1 was *very liberal*, and 7 was *very conservative*. Mean rating on this scale is 3.19 ($SD = 1.28$, $r.r. = 95.6\%$).

Higher numbers for political perspective represent political conservatism. Politics is significantly related to all supervariables except personal characteristics and has the strongest correlation among all demographic and background variables with the remaining three. Political conservatism is associated with traditional views about the validity and usefulness of intelligence tests and with low levels of bias and test misuse.

Many of the correlations in Table 5 are highly significant, but few of them are large. Other background

variables not shown, such as ethnic background, childhood family income, and having served as a news media source, show only very low correlations ($< .10$ and $> -.10$) with supervariables. Some attenuation of correlations resulted from the inclusion of missing values in supervariable creation. It should be noted, however, that the effects of demographic and background variables were also examined for each of the substantive questions separately using only nonmissing values, and the correlations were not substantially larger. Furthermore, the supervariables used in these analyses were formed from factors accounting for a relatively small amount of the data variance (39% total). These factors therefore do not represent strong patterns of responding, and one might expect the supervariables based on them to be resistant to prediction.

Stepwise multiple regression analyses were also performed with each of the supervariables as dependent variables and the demographic and background variables as predictors. Not surprisingly, in light of the data in Table 5, none of the regression analyses accounted for more than 19% of the variance in any of the supervariables.

General Discussion

What the foregoing results make clear is that those with expertise in areas related to intelligence testing hold generally positive attitudes about the validity and usefulness of intelligence and aptitude tests. These experts believe that such tests adequately measure most important elements of intelligence. Intelligence, as measured by intelligence tests, is seen as important to success in our society. Both within and between-group differences in test scores are believed to reflect significant genetic differences. There is support for the continued use of tests at their present level in elementary and secondary schools and in admissions to schools of higher education.

The picture that emerges from this survey is not wholly positive, however. Our sample of experts perceive problems with the influence of nonintellectual factors on test performance both within and between groups and particularly with certain test use practices. Intelligence and aptitude tests are seen as somewhat racially and socioeconomically biased. There is a widespread belief in frequent misinterpretation and overreliance on test scores in elementary and secondary schools, yet psychologists and educational specialists are generally in favor of the continued use of intelligence and aptitude tests in schools. Apparently, difficulties with bias and test use in the schools are not felt to be of sufficient magnitude to warrant an overall curtailment of otherwise useful decision-making tools. Respondents, as a whole, favor the decreased use of intelligence tests in employment.

One of the more puzzling aspects of our results is the relative lack of effect of within-sample variability in expertise. Our sample seems to vary rather widely in expertise, at least as measured by authorship, research, and academic specialty. The sample ranged from emeritus professors in the APA Division of Evaluation and Measurement with over 100 articles and chapters written on a broad range of testing issues to members of the Amer-

Table 5
Correlations Between Supervariables and Demographic and Background Variables

Variable	Test usefulness	Test bias	Personal characteristics	Test misuse
Gender ^a	-.15**	.22**	.18**	.15**
Age	.30**	-.10	-.06	-.12*
Authorship	.15**	-.06	-.13**	-.07
Political Perspective ^b	.31**	-.38**	-.06	-.17**

^a 1 = male, 2 = female. ^b Higher numbers correspond to conservatism.
* $p < .01$. ** $p < .001$.

ican Sociological Association with no measured experience in testing. Some of the diminished effect of expertise can be attributed to self-selection, as outlined earlier. It is also possible that expertise simply is not a major factor in opinions about testing.

Our inability successfully to predict differences in expert opinions about intelligence and testing on the basis of political and social attitudes is an even more interesting finding. It seems clear that despite the highly political climate surrounding testing, political ideology does not have a large influence on expert opinion. That political perspective accounts for less than 10% of the data variance and that experts hold generally pro-testing attitudes despite being slightly left of center politically are important points and must be contrasted with the heavy political influence apparent in public discussion about intelligence and aptitude testing. The relative immunity of expert opinion about testing to political influence, coupled with experts' knowledge of the empirical literature and firsthand experience, makes it imperative that the expert voice be heard in the public arena, particularly where important decisions are being made. Political decisions that have an impact on the lives of almost every member of society, as those about intelligence and aptitude testing do, need not be made entirely, or even primarily, on coldly rational grounds, but they must be informed decisions.

REFERENCES

- Bersoff, D. N. (1981). Testing and the law. *American Psychologist, 36*, 1047-1056.
- Block, N. J., & Dworkin, G. (Eds.). (1976). *The IQ controversy*. New York: Pantheon.
- Brim, O. G., Jr., Glass, D. C., Neulinger, J., & Firestone, I. J. (1969). *American beliefs and attitudes about intelligence*. New York: Russell Sage Foundation.
- CBS News. (1975, April 22). *The IQ myth* [Videotape]. New York: CBS, Inc.
- Cleary, T. A., Humphreys, L. G., Kendrick, S. A., & Wesman, A. (1975). Educational uses of tests with disadvantaged children. *American Psychologist, 30*, 15-41.
- Cronbach, L. J. (1975). Five decades of public controversy over mental testing. *American Psychologist, 30*, 1-14.
- Gould, S. J. (1981). *The mismeasure of man*. New York: Norton.
- Haney, W. (1981). Validity, vaudeville, and values: A short history of social concerns over standardized testing. *American Psychologist, 36*, 1021-1034.
- Herrnstein, R. J. (1971, September). I.Q. *Atlantic Monthly*, pp. 43-64.
- Herrnstein, R. J. (1982, August). IQ testing and the media. *Atlantic Monthly*, pp. 68-74.
- Houts, P. L. (Ed.). (1977). *The myth of measurability*. New York: Hart Publishing.
- Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.
- Kamin, L. J. (1974). *The science and politics of IQ*. Potomac, MD: Erlbaum.
- Larry P. v. Wilson Riles, 495 F. Supp. 926 (N. D. Cal. 1979).
- Lerner, B. (1980). The war on testing: David, Goliath & Gallup. *Public Interest, 60*, 119-147.
- Lerner, B. (1981). Representative democracy, "men of zeal," and testing legislation. *American Psychologist, 36*, 270-275.
- Lewontin, R. C., Rose, S., & Kamin, L. J. (1984). *Not in our genes*. New York: Pantheon.
- Owen, D. (1985). *None of the above: Behind the myth of scholastic aptitude*. Boston: Houghton Mifflin.
- Rothman, S., & Lichter, S. R. (1984). Personality, ideology and world view: A comparison of media and business elites. *British Journal of Political Science, 15*, 29-44.
- Sattler, J. M. (1982). *Assessment of children's intelligence and special abilities* (2nd ed.). Boston: Allyn & Bacon.
- Scarr, S., & Carter-Saltzman, L. (1982). Genetics and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 792-896). Cambridge, England: Cambridge University Press.