J. Vanecek, Eds. (Pergamon Press, New York, 1965), p. 215.
31. P. Nélis, *Compt. Rend. Soc. Biol.* **117**, 756 (1934).
32. D. A. Brown, M. C. Casewell, J. P. Quilliam, in *Recent Advances in the Pharmacology of Toxins*, H. W. Raudonat and J. Vanecek, Eds. (Pergamon Press, New York, 1965), p. 207.
33. R. Haque, *J. Bacteriol.* **93**, 525 (1967).
34. J. K. McClatchy and E. D. Rosenblum, *ibid.* **92**, 575 (1966); *ibid.*, p. 580.
35. G. P. Gladstone, *Postepy Mikrobiologii* **5**, 145 (1966).
36. I. Ginsburg and T. N. Harris, *Ergeb. Mikrob. Imm. Exp. Therap.* **38**, 198 (1964).
37. For a contrary view, see F. Bergmann, S. Leon, M. Chaimovitz, F. Benzakein, *Brit. J. Pharmacol. Chemotherap.* **17**, 159 (1961).
38. A. W. Bernheimer and L. L. Schwartz, *J. Bacteriol.* **87**, 1100 (1964).
39. W. Wilbrandt, *Arch. Ges. Physiol.* **245**, 22 (1941).
40. A. W. Bernheimer, *J. Gen. Physiol.* **30**, 357 (1947).
41. G. Hirsch, A. W. Bernheimer, G. Weissmann, *J. Exp. Med.* **118**, 223 (1963).
42. D. Zucker-Franklin, *Amer. J. Pathol.* **47**, 419 (1965); A. M. Woodin and A. A. Wieneke, *Exp. Cell Res.* **43**, 319 (1966).
43. M. G. MacFarlane and B. C. J. G. Knight, *Biochem. J.* **35**, 884 (1941).
44. M. G. MacFarlane, *Mechanisms of Microbial Pathogenicity* (Cambridge Univ. Press, Cambridge, 1955), p. 57.
45. H. M. Doery, B. J. Magnusson, J. Gulasekharam, J. E. Pearson, *J. Gen. Microbiol.* **40**, 283 (1965).
46. G. M. Wiseman and J. D. Caird, *Can. J. Microbiol.* **13**, 369 (1967).
47. W. R. Chesbro, F. P. Heydrik, R. Martineau, G. N. Perkins, *J. Bacteriol.* **89**, 378 (1965).
48. A. W. Bernheimer, in *Streptococcal Infections*, M. McCarty, Ed. (Columbia Univ. Press, New York, 1954), p. 19.
49. J. G. Howard, K. R. Wallace, G. P. Wright, *Brit. J. Exp. Pathol.* **34**, 174 (1953).
50. A. W. Bernheimer, *J. Bacteriol.* **91**, 1677 (1966).
51. ——— and M. Davidson, *Science* **148**, 1229 (1965).
52. K. Hirschhorn, R. R. Schreibman, S. Verbo, R. H. Gruskin, *Proc. Nat. Acad. Sci. U.S.* **52**, 1151 (1964).
53. G. Weissmann, B. Becher, G. Wiedermann, A. W. Bernheimer, *Amer. J. Pathol.* **46**, 129 (1965).
54. C. M. MacLeod, in *Cellular and Humoral Aspects of the Hypersensitivity States*, H. S. Lawrence, Ed. (Hoeber-Harper, New York, 1959), p. 621.
55. A. W. Bernheimer and L. L. Schwartz, *J. Pathol. Bacteriol.* **89**, 209 (1965).
56. L. F. Hewitt and E. W. Todd, *ibid.* **49**, 45 (1939).
57. I. Ginsburg, N. Elias, M. Heller, *Biol. Conf. Oholo* **10**, 64 (1965).
58. A. D. Bangham, M. M. Standish, G. Weissmann, *J. Mol. Biol.* **13**, 253 (1965).
59. J. Robinson, F. S. Thatcher, J. Montford, *Can. J. Microbiol.* **6**, 183 (1960).
60. B. J. Magnuson, H. M. Doery, J. Gulasekharam, *Nature* **196**, 270 (1962).
61. G. Weissmann, G. Sessa, A. W. Bernheimer, *Science* **154**, 772 (1966).
62. J. Meduski and M. Wolkowa, *Med. Doswiadczalna i Mikrobiol.* **10**, 246 (1958).
63. J. E. Alouf and M. Raynaud, *Nature* **196**, 374 (1962).
64. A. Kayser, *Comp. Rend. Acad. Sci. Ser. D*, **263**, 693 (1966).
65. A. Yoshida, *Biochim. Biophys. Acta* **71**, 544 (1963).
66. G. P. Gladstone and W. E. van Heyningen, *Brit. J. Exp. Pathol.* **38**, 125 (1957).
67. M. S. Artenstein, M. A. Madoff, L. Weinstein, *Yale J. Biol. Med.* **35**, 373 (1963).
68. A. W. Bernheimer and L. L. Schwartz, *J. Bacteriol.* **89**, 1387 (1965).
69. A. W. Bernheimer, *J. Protozool.* **10**, 166 (1963).
70. H. Keiser, G. Weissmann, A. W. Bernheimer, *J. Cell Biol.* **22**, 101 (1964).
71. G. Weissmann, H. Keiser, A. W. Bernheimer, *J. Exp. Med.* **118**, 205 (1963).
72. A. W. Bernheimer, unpublished data.
73. R. M. Fauve, J. E. Alouf, A. Delaunay, M. Raynaud, *J. Bacteriol.* **92**, 1150 (1966).
74. J. Jeljaszewicz, S. Szmigielski, M. Korbecki, C. Zak, *J. Infect. Dis.* **115**, 421 (1965).
75. A. Kayser and M. Raynaud, *Ann. Inst. Pasteur.* **108**, 215 (1965).
76. G. P. Gladstone and A. Yoshida, *Brit. J. Exp. Pathol.* **48**, 11 (1967).
77. H. Ikezawa and R. Murata, *J. Biochem. Japan* **53**, 217 (1964).
78. M. G. MacFarlane and N. Datta, *Brit. J. Exp. Pathol.* **35**, 191 (1954).
79. Aided by grants from the Life Insurance Medical Research Fund, PHS grant (AI-02874-09), and by a PHS research career program award (5K6-AI-14, 198-06).

Standardized Ability Tests and Testing

Major issues and the validity of current criticisms of tests are discussed.

D. A. Goslin

Standardized ability tests have been a source of considerable controversy in recent years. Growing competition for jobs and for all educational opportunities has intensified the search for better ways to evaluate individual abilities and aptitudes and to identify intellectual potential at progressively earlier ages. Standardized tests of various types increasingly are used to identify applicants throughout the educational system, as well as by the military, the civil service, and business and industry (*1–3*).

This reliance on results of standardized tests has caused questions to be raised about the validity of the tests used, as well as their effects on those who take them and on the society that uses them to differentiate among its members. Thus far, there have been very few, if any, attempts to bring together all of the criticisms that have been leveled against tests, and to place them in an analytical framework that would permit a systematic evaluation of their validity. In this paper the validity of standardized tests is discussed, and major criticisms of tests are summarized within such a framework.

The author is staff sociologist at Russell Sage Foundation in New York City. This paper was prepared originally for the Commission on Testing of the College Entrance Examination Board as a background report for the Commission's meeting, 5–6 June 1967, Washington, D.C. The author gratefully acknowledges the permission of the College Board to publish it in *Science*.

Three Variables of Criticism

Criticisms of testing relate to three variables: the type of test, how it is used, and assumptions regarding its validity. First, the type of test being used must be considered. Ability tests may be divided into tests that attempt to measure inherent capabilities, potentials, or abilities acquired over a long time, and tests designed to measure specific achievements.

Intelligence and aptitude tests are implicitly assumed to measure a relatively deep and enduring quality. This quality may be viewed as changeable; however, startling changes are assumed to be rare except under specific conditions, as when extreme cultural deprivation is ameliorated. Intelligence and aptitude tests therefore generate anxiety in people tested. The high cultural value placed on intellectual abilities in our society also makes any instrument which purports to measure general intellectual abilities a source of fascination. For these reasons, such tests have been a major source of controversy and debate.

Although less often perceived as unfair, since they measure skills acquired in a particular area over a short time, achievement tests potentially exert a considerable influence on subject matter and teaching methods, as well as on what skills appear desirable. Among all tests, they are distinctive in that it is

easier in the case of an achievement test to see what one is measuring, since the universe of abilities being sampled by the test is theoretically finite and far more easily specified.

The second variable is the use to which the test is put. Test results may be used for selection and placement, or counseling, and sometimes both. A test used to select among a group of candidates for a job, or among applicants for admission to a school, or a test used to assign individuals to specific groups (like tracks in a school) has an essentially predictive function. It is used to predict individual performance with respect to that of the other members of the group.

Tests may also be used as a basis for providing an individual with information about his abilities and aptitudes. This use of tests is theoretically different from that previously mentioned because the information provided to the "counselee" is intended to enable him to decide about his future. In the former case, although the individual sometimes decides for himself (as whether or not to apply), others ultimately decide for him. However, counseling frequently directs the individual to one of several alternatives. In this case, depending on the information and how it is transmitted, the counselor may actually be the decision-maker.

Finally, criticisms may either question the validity of tests or they may have little or nothing to do with whether the test measures what it is supposed to measure. Here we must ask: Is the force of the criticism affected by whether we assume the test to be a valid measure of what it is supposed to measure, or not?

Criticisms of the Validity of Ability Tests

Several critics have claimed that certain characteristics of tests make them unfair and invalid predicators for certain individuals or groups. These critics have singled out three types of individuals.

First, Hoffmann (*4*) and others have claimed that these tests are unfair to deep thinkers. Critics who take this position claim that certain items on standardized tests penalize bright students because they are ambiguously worded or because the alternatives presented include one or more options (scored as incorrect) that the mediocre student passes by, but which the extremely bright student correctly perceives as being possibly correct answers. One cannot dispute the fact that Hoffmann and others have demonstrated clearly the existence of such items on tests that are currently in use.

Although it is not known whether any extremely bright students have actually suffered because of poorly written tests, Hoffmann's point is valid, at least in the abstract. It seems unlikely, however, in our achievement-oriented society, that very many geniuses remain undiscovered, regardless of their performance on standardized tests (or more important, that more geniuses are missed because of standardized tests than would be missed with alternative selection techniques).

Second, any test designed to be given to individuals in our heterogeneous society will discriminate against people with a cultural background different from that of the majority. To take an extreme case, if a pupil cannot read English because Spanish is spoken at home, he is not likely to do well on tests of reading comprehension in English, or, in fact, on any test written in English. Members of any group whose life experiences differ significantly from those on which the test was standardized will also be at a disadvantage.

Partly, this is a problem of standardization. Conceivably, special norms could be developed on any test for every distinctive group likely to take the test, so that both inter- and intragroup comparisons could be made. But another principle is involved. Most standardized tests are designed to predict success of individuals in the broader society, or in the setting in which the individual wishes to gain admission. Thus, tests are doing their job when they discriminate. If facility in English is assumed necessary for success, then a test of that facility is not unfair. In such cases, it can be pointed out that it is not the test which is unfair, but rather the circumstances which have permitted the deprivation to persist. However, any inferences about the general intellectual abilities of members of disadvantaged or other special groups based on test scores should be avoided at early ages.

Finally, tests may be unfair to individuals who lack special skills required for taking standardized tests. For almost everyone, these skills may be assumed to develop as a result of repeated contact with tests. Some individuals, however, take more tests than others. The amount of experience required to make this factor an unimportant influence in test performance is unknown. It may be assumed, however, that tests are unfair to individuals without the requisite experience with tests (*5*). Hence, extensive testing in elementary and junior high school is beneficial, but inequalities may be created when some schools test more frequently than others. The problem is acute for foreigners from countries where tests are not widely used (for example, foreign applicants to American graduate and professional schools).

Factors Affecting Validity of Tests

Standardized ability tests are not perfect predictors of subsequent performance, even in situations that require abilities similar to those required on the test. Highest coefficients of correlation between test scores and measures of subsequent performance are obtained for short-range academic performance (*6*). For example, scores of standardized tests given in the 12th grade predict first-year college grades fairly well. As the length of time between the test and criterion situation increases, the magnitude of the correlation is reduced. Similarly, as the criterion situation becomes more dissimilar from the test situation, the correlation is reduced. Thus, most existing studies show no correlation between test scores and subsequent occupational success (nor is any correlation shown between academic performance as measured by grades and subsequent occupational success). Because test scores correlate only moderately with long-range academic performance and not at all with postacademic performance, one can raise serious questions about their usefulness and reliability.

Three factors contribute to this lack of correlation. First, it is often difficult to establish clear criteria for successful postacademic performance. Many studies have relied on performance ratings by professional colleagues, fellow workers, or superiors. These are frequently unreliable and are based on other factors, such as personal qualities. Use of objective criteria, such as number of scientific papers published, may be criticized as being superficial.

Second, there is the problem of range restriction. Accurate predictions about the relative performance of individuals are easily made where there are sizable differences between individuals; a high degree of variance in the distribution of abilities measured makes prediction easy. However, where differences among members of the group tested are

small, it is difficult to predict later performance of the members of the group relative to one another. Thus, predictions in a homogeneous group, such as college graduates, are risky. The phenomenon of range restriction accounts in large part for the lack of correlation between either test scores or academic performance and occupational success among able students.

Third, one should not assume that there is a linear relation between qualities measured by a standardized test and occupational success. The assumption that intelligence alone determines success is superficial. In fact, many studies have revealed that the relation between intellectual abilities and success in our society is very complicated. For example, although Terman demonstrated clearly that his gifted group as a whole was more successful than less intellectually able groups, he found no relationship between intelligence and later performance within the gifted group (*7*).

These findings are corroborated by the previously noted lack of correlation between college performance and subsequent nonacademic success and suggest that intellectual abilities may function as a threshold variable in relation to occupational advancement. A minimum level of intelligence is obviously required for most occupations, but once at or above this threshold, individual achievement relative to others in the same field is determined by qualities not measured by tests of intellectual abilities.

It should be noted that fields of endeavor differ not only in basic requirements of intelligence, but also with respect to the amount of difference made by increments over this level in one's chances of achieving success. In other words, qualities other than basic intelligence are more important in some fields than in others. [Incidentally, this does not have to be the case; it just happens that our society works this way at present. One could, for example, imagine a society in which a perfect correlation between intelligence and success could be achieved by assigning all jobs and status in the society on the sole basis of intelligence (*8*).]

There is some controversy about whether ability tests measure innate capabilities (presumed unchangeable) or learning. Few people with any sophistication in psychometrics believe that even intelligence tests measure only innate capabilities. However, there are significant differences in opinion about whether the qualities measured by intelligence tests are more or less influenced by learning than by inherent potential. Assumptions about exactly what the test measures are likely to have an important effect on how test scores are used. If one interprets a child's performance as an indication of what he has learned (as opposed to a result of innate capability), then one is less likely to make long-run predictions about the child's ultimate success on the basis of his test scores (for example, his motivation might increase, and he might do better next time).

One of the most important criticisms of tests is that they contribute to their own validity by functioning as self-fulfilling prophecies. Hypothetically, a child who does well on a test, and, as a consequence of his performance, is placed in an advanced class, or receives special attention from his teachers, or who is admitted to a good university, is more likely to do well than the one whose score was lower. The likelihood that the optimistic prediction made on the basis of a high test score will be fulfilled is therefore increased because the person who scores high receives special advantages, whereas the individual who does poorly is often denied opportunities.

Experimental data from a recent study by Rosenthal and Jacobsen (*9*) confirm this hypothesis. They gave all of the children in four California elementary schools an ordinary intelligence test at the beginning of the school year. They informed the teachers that the test they had given was specially designed to identify children who could be expected to show substantial I.Q. gains during the coming year. In each class, they then selected at random ten children and informed the teachers that these children had done particularly well on the test. This group in each class formed the experimental group, and the remainder of the children in each class served as the control group. An intelligence test given at the end of the school year showed that the experimental groups in grades kindergarten, one, two, and three had made significant gains in I.Q. when compared to the children in the control groups. In addition, teachers rated children in the experimental groups as being superior to those in the control groups in personal qualities, such as cooperativeness, interest in school affairs, and social adjustment. These data reveal that teachers' expectations contributed substantially to the increased test scores of the children in the experimental groups. Here the first test score reported to the teachers became a self-fulfilling prophecy. The implications of this point are far-reaching, especially for policies concerning the use of standardized intelligence tests in the elementary grades.

Criticisms Independent of the Validity of Tests

The following criticisms may be hypothesized to hold, whether one argues that tests are valid measures of ability or not. In some cases, the force of the criticism is increased if one assumes tests to be highly valid predictors. These criticisms, therefore, stem from the potential social effects of testing, rather than from questions regarding the accuracy of tests.

Standardized ability tests are used throughout the educational system, and children take such tests at periodic intervals. In addition, the spread of the technology of standardized test construction has led many teachers to make use of objective questions in tests they construct. It has been suggested that continual exposure to multiple-choice items during the elementary and secondary grades tends to result in constriction of children's ability to reason. In particular, it is claimed that emphasis on evaluation techniques in which there is always a right and wrong answer makes it difficult for children to deal with issues on which there is no clear right or wrong answer (*10*). Children, it is claimed, are therefore handicapped when they attempt to work through questions involving ethical or philosophical judgments, or when arriving at a decision depends upon identifying the assumptions one is going to begin with.

There is no proof that this is a valid criticism. Colleges claim that incoming students do not write as well as formerly, but there is no way of knowing if the older generation is just complaining about the new one, or, if true, whether it is because proportionally more people are going to college today. Lack of a suitable control group (that is, college students who have not taken standardized tests) makes research on this topic difficult.

When a student takes a college entrance examination or almost any standardized test, not only he, but also his teachers and his school, are being tested, since his performance reflects his training. As a consequence, tests have a potentially significant impact on subject

matter and teaching methods. Only a very small minority of teachers interviewed in a recent study (*3*) claimed to spend much time preparing students for standardized tests or indicated that they have ever altered a course because the subject matter covered by a standardized test was different from what they normally taught. Nevertheless, there is some evidence that in many situations standardized tests do exert an influence on what is taught. The well-known New York Regents' examination program is pertinent here. Since both teachers and schools were being evaluated along with students, there was, and still is, considerable pressure to prepare students to take the Regents' achievement examinations. Reports of students being drilled on old copies of the Regents' examinations were common. That tests have had an impact on the curricula in this case cannot be disputed (*11*).

Whether or not teachers make special efforts to prepare students for taking particular standardized tests, such tests can have a more general impact on curricula. For example, widely used external examinations, like the College Board achievement tests, may result in pressure on a school system to adopt a new curriculum if the school perceives that the content covered by the test differs significantly from that which is being presented in the school. Thus, standardized tests based on the new mathematics curriculum can be expected to speed the adoption of this curriculum in schools.

It should be noted that such an effect is not necessarily deleterious. Standardized tests may raise school standards as often as they limit innovation and experimentation. This, of course, was the idea behind the Regents' examination program when it was initiated. The problem is striking a balance between raising standards and setting arbitrary limits.

More and more schools, colleges, and testing agencies are giving individuals either their specific score or percentile rank, or a general idea of how they did on standardized tests. Regardless of how such information is transmitted to the examinee, it may be hypothesized that it will affect self-image, motivation, and aspirations in some cases. Users of tests have alternately been criticized for withholding test scores and for indiscriminately giving results. The effect of receiving information about one's abilities will depend on the perceived legitimacy of the source of the information (for example, the pupil's counselor), the perceived accuracy of the test, and the degree to which the test score confirms one's own estimate, including how threatening or rewarding it is. Obviously, individuals make use of many different types of information in arriving at an estimate of their abilities; standardized test scores are only one of many ways in which individuals get information about their capabilities. Data from a national sample of high school students (*2*) indicate that test scores are of relatively minor importance in shaping self-estimates of ability, in comparison with such things as school grades, comments of peers and parents, and contact with teachers.

Test scores do have a potentially great impact when the individual's self-estimate varies considerably from his test score and when he cannot rationalize his poor performance, or when the score is substantially higher than his estimate. Under such conditions we may expect a shift in self-estimate of ability to affect the individual's aspirations, his motivation, and, secondarily, future personal decisions. We should also consider the consequences for overall aspiration levels in the society of a system in which individuals are classified very early with respect to their abilities and available opportunities for the future.

The use of any single criterion or set of criteria to sort individuals into groups or to decide which individuals will be admitted to a group affects the structure and characteristics of groups so formed. These implications may be examined under the following headings: (i) social structure within groups, (ii) tendencies toward uniformity in the characteristics of group members, and (iii) implications for the society as a whole.

With regard to (i), the current widespread use of standardized tests to allocate students to instructional groups or to tracks within schools causes social differentiation within schools based on qualities measured by standardized tests. Ability groupings reduce social contact between pupils of differing levels of ability (as measured by standardized tests). Research indicates that such differentiation may affect performance levels of low-ability pupils negatively, while not significantly facilitating the performance of high-ability pupils (*12*). In addition, it is clear that ability grouping impedes the process of acculturation of members of culturally deprived groups, who tend to end up together in the low-ability groups.

As for (ii), the use of any single criterion for forming groups produces a strong tendency toward uniformity in the members of the group. Our elite colleges and universities, for example, have difficulty achieving diversity in the student body while admitting only students of exceptional ability. The problem becomes more acute when standardized tests are heavily relied upon as a measure of intellectual ability.

Concerning (iii), Wolfle has pointed out that the success of modern, complex societies depends in large part on the availability of a talent pool in which a great diversity of abilities and skills is represented (*13*). To create such a talent pool, rewards of social status, prestige, and economic returns must be provided for individuals possessing many different talents. A tendency to rely heavily on standardized tests of a limited set of intellectual skills in the allocation of opportunities for achievement must necessarily reduce the diversity of talent available. Here, we might consider testing more abilities than those measured by current tests. We must also ensure adequate rewards for individuals possessing abilities not measured by tests, but which are important for the successful functioning of the society.

Do Tests Invade Privacy?

A test is a potential invasion of privacy because personal information is made available to others. Very important values in American society suggest that individuals have the right to decide to whom and under what conditions they will make available to others information about themselves. Correlative to this point, however, is the fact that participation in the society carries with it certain obligations and responsibilities. Further, certain groups clearly have the right to demand information from those who want the privileges of group membership. Thus, no one is likely to object to being given a driving test before being permitted to operate a motor vehicle. Similarly, few people object to the requirement that they must take an entrance test in order to gain admission to a university or college. In each case, the right of a group to information that is relevant to the stated objectives and goals of the group has been established.

Two important questions remain,

however. First, under what conditions does a group have the right to ask aspiring members for information that is irrelevant to the purposes and goals of the group? In order to answer this question, it is probably necessary to make a distinction between public and private groups. A private group usually has the right to ask of applicants for membership anything, whether relevant or irrelevant. The applicant then decides whether he wishes to reveal this information. In the case of a group supported by the society as a whole, including all of the potential applicants to the group, the situation is more difficult. Would it be, for example, legitimate for the state to ask individuals to reveal information about their sexual behavior as a requirement for obtaining a driver's license? Most of us would, I think, object to such a requirement on the grounds that it represents an invasion of our privacy that is not justified by the service being rendered. The issue is one of relevance: must the school have such information in order to do its job?

There is, however, a second and more difficult problem in the case of school testing. In each of the cases presented above, the individual retains a choice as to whether he will submit himself to the test or not. Thus, if an individual does not want to take the College Board Scholastic Aptitude Test (SAT), he does not have to. Nor does he have to submit to a driver's test. As a result of his decision, he may have to give up his chances of attending certain colleges or driving an automobile, but the choice in each case is his. But, for the most part, a child does not have a choice about whether he will take tests or not, including standardized tests. A parent might move to another community, in which the school system did not use standardized tests (if he could find one), or he might send his children to a private school that did not administer tests (if he could afford one). For most parents these are not realistic alternatives.

Does this constitute an invasion of privacy? Carried to its extreme, an affirmative answer leads one to the conclusion that children should be permitted to refuse to take all tests, even those given by their teachers in class. Although this sounds absurd, it is not an unreasonable claim. If a child refused to participate in classroom tests, he would fail his courses and would not be promoted, but this would be his (or his parents') decision. The school clearly does have a right to require pupils to demonstrate their proficiency in school subjects before according them advanced status. But does the school also have the right to require pupils to demonstrate their general intellectual ability apart from their proficiency in specific subjects? If a child refused to take an I.Q. test given in school, would he fail his course? Does a school need such information in order to decide whether or not a child should be promoted?

If one concludes that a school has the right to collect information about intellectual abilities of its pupils, does the school also have the right to withhold this information from the pupil and his parents? Conversely, what right do parents and pupils have to know what information the school possesses about them? In at least one case (in New York State), the courts ruled that parents do have the right of access to information on the pupil's permanent record card maintained by the school (*14*).

Summary and Conclusions

At the outset a distinction was made between criticisms directed at the validity of tests and criticisms not affected by the validity of the tests. It was noted further that all criticisms of tests must take into consideration the type of test and the use to which the test is put.

Criticisms of the validity of tests involved the following issues: (i) tests may be unfair to certain groups and individuals, including the extremely gifted, the culturally disadvantaged, and those who lack experience in taking tests; (ii) tests are not perfect predictors of subsequent performance; (iii) tests may be used in overly rigid ways; (iv) tests may not measure inherent qualities of individuals; and (v) tests may contribute to their own predictive validity by serving as self-fulfilling prophecies.

Criticisms that are more or less independent of test validity included the effects of tests on (i) thinking patterns of those tested frequently; (ii) school curricula; (iii) self-image, motivation, and aspirations; (iv) groups using tests as a criterion for selection or allocation, or both; and (v) privacy. Several concluding remarks are in order:

1) This paper has focused almost entirely on *criticisms* of tests. However, the positive value of standardized tests should not be ignored. Here we must keep in mind what possible alternative measures would be used if standardized tests were abandoned.

2) We must begin thinking about tests in a much broader perspective—one that includes consideration of the social effects of tests as well as their validity and reliability.

3) Finally, an effort should be made to develop rational and systematic policies on the use of tests with the culturally disadvantaged, the dissemination of test results, and the problem of invasion of privacy. Such policies can be formulated only if we are willing to take a long hard look at the role we want testing to play in the society. Standardized tests currently are a cornerstone in the edifice of stratification in American society. It is up to the social scientist to conduct research that will enable policy makers in education, business and industry, and government to determine in a consistent and rational way the ultimate shape of this edifice.

References and Notes

1. A number of surveys of the extent of testing and test use have appeared in recent years. See, for example, D. A. Goslin, *The Search for Ability* (Russell Sage Foundation, New York, 1963).
2. O. G. Brim, Jr., D. A. Goslin, D. C. Glass, I. Goldberg, *The Use of Standardized Ability Tests in American Secondary Schools* (Russell Sage Foundation, New York, 1964).
3. D. A. Goslin, R. R. Epstein, B. Hallock, *Testing in Elementary Schools* (Russell Sage Foundation, New York, 1965); D. A. Goslin, *Teachers and Testing* (Russell Sage Foundation, New York, 1967).
4. B. Hoffmann, *The Tyranny of Testing* (Crowell-Collier, New York, 1962).
5. Numerous studies have examined the effects of "practice" or "coaching," or both, on test scores. Most have found little or no significant effects. A distinction should be made between the effects of specific practice or coaching experiences and the effects of long-term exposure to standardized tests.
6. For a summary of research on the prediction of academic performance, see D. E. Lavin, *The Prediction of Academic Performance* (Russell Sage Foundation, New York, 1965).
7. L. M. Terman, *The Gifted Group at Mid-Life* (Stanford Univ. Press, Stanford, 1959).
8. Such a society is envisioned by Michael Young in his book, *The Rise of the Meritocracy* (Thames and Hudson, London, 1958).
9. R. Rosenthal and L. Jacobsen, *Pygmalion in the Classroom: Self-Fulfilling Prophecies and Teacher Expectations* (Holt, Rinehart and Winston, New York, in press).
10. Often, multiple choice tests require that the "best" answer from the alternatives provided be selected. This choice is, by definition, the "right" answer. Although such tests may measure judgment effectively, they still may cause the person who takes tests to see the world in "right" or "wrong" terms.
11. H. M. Brickell, *Organizing New York State for Educational Change: A Report to the State Commissioner of Education* (State Department of Education, Albany, N.Y., 1962).
12. A. Yates, *Grouping in Education* (Wiley, New York, 1966).
13. D. L. Wolfle, "Diversity of Talent," *Amer. Psychol.* **15**, 539 (1966).
14. *Van Allen v. McCleary* 27 Misc 2d 81, 211 NYS 2d 501 (Superior Court, Nassau County, 1961).