

The Social and Educational Consequences of Standardized Testing:

A point of view

Abstract

Standardized testing has figured prominently in Canadian education from the 1920s to the present, affecting the life chances of Canadians for education, jobs, and happiness. This paper examines the historical and sociological foundations of the testing movement and the consequences for the people whose lives are affected by the tests they take.

The curmudgeon of American print journalism, H.L. Mencken, was fond of saying that "no one ever went broke underestimating the intelligence of the American public." With testing a 300-400 million dollar industry, Mencken was certainly correct. It is nevertheless the case that testing is more than a lucrative industry. Standardized tests are administered and interpreted within a particular social frame of reference. The consequences of how one scores on tests affect one's life chances -- for education, for jobs, and for happiness. This paper will examine the historical and sociological foundations of the testing movement and the consequences for the people whose lives are affected by the tests they take.

Historical and sociological foundations

In less than one hundred years, standardized testing has grown from infancy into an enormously profitable and influential industry. The rapid growth of the testing movement and its general acceptance were related to a number of developments. These included the rapid expansion and transformation of the

economy, the gradual sophistication of testing methods, the belief in innate intelligence, and the expansion of educational systems and their functions. The change from an agricultural economy to a manufacturing economy deeply affected our conception of human worth. The shift of economic power from the countryside to the city and the emergence of an achievement-oriented social system permitted the development of a set of social beliefs which could not have existed in an agrarian society based upon lineage. In agrarian societies, attempts to change one's social position were inconsistent with the idea of a social hierarchy based upon birth. Ability, ambition, and effort did not affect one's position in an agrarian society. With industrialization, birth gave way to achievement as the basis for locating people in the social structure. This transformation provided the social context for the development of methods for testing achievement.

Pioneer theories. During the latter part of the 19th century, influential people such as Herbert Spencer (1820-1903), William Graham Sumner (1840-1910), and Charles Darwin's cousin, Francis Galton (1822-1911), were applying ideas from the biological and physical sciences to society. Spencer, who coined the term "survival of the fittest," believed that homogeneity was inherently unstable. He reasoned that the process of increasing social differentiation would eventually result in "the establishment of the greatest perfection and the most complete happiness." According to historian Richard Hofstadter, Spencer held the view that:

...the pressure of subsistence upon population must have a beneficent effect upon the human race. This pressure had been the immediate cause of progress from earliest times. By placing a premium upon skill, intelligence, self-control and the power to adapt through technological innovation, it had stimulated human advancement and selected the best of each generation for survival. (Hofstadter, 1969, p.39)

According to Spencer, social theory should recognize the selection principle from biology and refrain from violating it by "the artificial preservation of those least able to take care of themselves" (Hofstadter, 1969, p.44).

The rapid development of the use of standardized tests of intelligence and ability depended in large part on their claim to be scientific and unbiased. Francis Galton believed that intelligence tests could be used to divide humanity on the basis of innate differences and thereby provide the means to control the production of "fine human offspring" (eugenics). Galton's **Hereditary Genius: An Inquiry into its Laws and Consequences** was published in London in 1869. Galton proposed to "...range men according to their natural abilities, putting them into classes separated by equal degrees of merit, and to show the relative

number of individuals included in the several classes..." (Galton, 1869, p.26). Galton coined the term "eugenics" to refer to:

...the science which deals with all influences that improve the inborn qualities of a race; also with those that develop them to their utmost advantage...the aim of Eugenics is to represent each class or sect by its best specimens; that done, to leave them to work out their own common civilizations in their own way. (Collini, 1979, pp.201-202)

Galton, who founded the Eugenic Society in 1870, had studied men of science, "genius", and the Fellows of the Royal Society. His investigations had persuaded him that nature was much more influential than nurture and that the upper social strata produced a larger number of talented individuals.

The development of the concept of "statistical correlation" by Galton's protege, Karl Pearson (1857-1936), and the enunciation of the idea of "factor analysis" by Charles Spearman furthered the belief that it was possible to design scientific, objective, and unbiased tests of innate intelligence (Martindale, 1960). Karl Pearson, Galton Professor of Eugenics at London University and Director of the Francis Galton Laboratory, argued that biological factors "...are dominant in the evolution of mankind; these and these alone, can throw light on the rise and fall of nations, on racial progress and national degeneracy." Pearson estimated that ninety percent of a person's capacity was determined by heredity. He was among those of the period who argued that education should vary for individuals and groups in relation to their abilities (Spearman, 1904).

First tests. By 1905, Alfred Binet and Theodore Simon had produced the first scale for the measurement of intelligence. When Galton's volume was issued in North America in 1914, it was followed by literally hundreds of studies of genius. In North America, Lewis Terman, Henry Goddard, and Robert Yerkes were the leaders of the testing movement. As early as 1912 Henry Goddard had begun to administer the Binet-Simon intelligence test to immigrants. Then, in 1916, Lewis Terman of Stanford University translated and adapted the Binet-Simon scale and published the first Stanford-Binet intelligence scale. This scale was revised in 1937 and again in 1960 and is still one of the most widely used tests. Finally, in 1917, Robert Yerkes of Harvard University developed the "Alpha" and "Beta" intelligence tests which were administered to two million American army recruits (Terman, 1916, 1937; Kamin, 1974).

Although Binet had been careful to state that the Binet-Simon scale measured competence and not a genetic quality, the Binet-Simon scale was soon labelled as a measure of innate intelligence. Without evidence, educational psychologists of the

period made a leap of faith to the position that general intelligence as measured by intelligence tests was an innate quality; Terman, Goddard, and Yerkes were among those who subscribed to this belief.

Impact of economics and war. In the first 20 years of this century, industry experienced a period of soaring production, rapid growth, and the concentration of capital. This combination of changes demanded an increase in the number of workers with specific productive, managerial, technical, and clerical skills. Corporate capitalism was faced with the problem of having to select suitable trainees who would provide the largest return on the investment involved in training. Intelligence, achievement, and vocational tests offered corporations an ideal way of selecting from the mass of available workers those trainees who would most efficiently "repay" the costs of training through their contributions to production.

The First World War increased the need for more educated and skilled labour organized into a pyramid-shaped structure. New tests helped to provide justification for this hierarchy. Because of the "scientific" stature accorded to the belief that intelligence was innate, workers could be classified on the basis of "intelligence" scores and taught to accept their location in a workplace that was structured inequitably. Psychological tests thus provided both the means by which workers could be efficiently selected for training and the rationale for the maintenance of an acquiescent workforce. Private corporations saw the development of these tests as a natural and beneficial part of running the economy.

Uses in education. Standardized tests of intelligence and ability were also considered scientific tools that could be used to classify people of different races and social classes in order to determine who should be educated. "Intelligence" and associated achievement tests were thought ideally suited to the rapidly expanding education systems which were geared to serve the industrial structure. These tests offered what seemed to be an unbiased scientific means of measuring students and maintaining the hierarchical education system.

The rise of universal education systems and the expansion of higher education in Britain and North America paralleled the vast economic changes taking place during the first two decades of the century. In Britain, the educational system was gradually extended largely as a result of pressure from members of the working class who were demanding better education for their children. In North America, working class pressure and the massive tide of immigrants helped bring about the expansion of the educational system. Intelligence and ability testing made it possible to select and stream the increased numbers of people who wished to have access to the educational system.

North American school systems were based on an ideology of equality. Schools and colleges used the new tests to classify students and to preserve the appearance of equal opportunity. The scientific status of these tests provided an easy and seemingly legitimate way to select students by ability. Educational reformers argued that these tests would allow schools and teachers to "fit education to the individual needs of students." Education had become a hierarchical meritocracy firmly supported by "objective and unbiased" test scores.

Nativistic viewpoints. Nativism, protection of the interests of native inhabitants against those of immigrants, is a persistent theme in the literature devoted to intelligence. For Canada's first comprehensive survey of a provincial school system, the Putman and Weir **Survey of the (British Columbia) School System**, Professor Peter Sandiford undertook an "...objective survey of the intelligence and school achievement of the pupils of the Province..." In a sub-section of his assessment entitled "The Mental Capacity of B.C. Students According to the Place of Birth," Sandiford wrote:

The object of this study was to determine the influence of immigration from various centres upon the general intellectual level of the B.C. community. If the people coming into the Province are of higher intelligence than those already in it, the general level will be raised. This, of course, is a result greatly to be desired. On the other hand, continued immigration of inferior stock can end only in disaster. (Putman and Weir, 1925, p.458)

A paragraph later, Sandiford elaborated upon the nativistic theme he had developed above:

The low intelligence of pupils born in Continental Europe and the rest of the world is the outstanding feature of this investigation. These high school pupils, normal school and university students have a median intelligence quotient lower than 100; that is, lower than that of the community as a whole. Immigration from these areas lowers the average intelligence of the population. (p.459)

Later in the report, Sandiford describes the mental capacity of the Japanese and Chinese pupils enrolled in Vancouver Schools. Recognizing that "...tests involving a use of English language would not be fair..." to Japanese and Chinese pupils, the testers used a battery of fifteen tests to assess the intelligence of 155 Chinese and 150 Japanese pupils. The subsequent discussion reveals a number of things. First, the nativistic viewpoint that characterized the earlier discussion of intelligence is unmistakable. Second, Sandiford contradicts his earlier expression of the

desirability of raising the general level of intelligence by immigration:

...it will be seen that the Japanese are superior to the Chinese and both are greatly superior to the average white population. The superiority is undoubtedly due to selection. In the main it is the Japanese and Chinese who possess the qualities of cleverness, resourcefulness, and courage who emigrate to British Columbia; the dullards and less enterprising are left behind. This superiority of emigrant stock is no new phenomenon in world history. There are those who maintain that Great Britain owes her eminent position in the world to the fact that only the clever and sturdy could secure a footing on her shores.... But from the political and economic standpoints the presence of an industrious, clever, and frugal alien group, capable so far as mentality is concerned, of competing successfully with the native whites in most occupations they mutually engage in, constitutes a problem which calls for the highest quality of statesmanship if it is to be solved satisfactorily. (p.508)

Acknowledging the "great superiority of the Japanese and Chinese alien groups over all others," and the fact that "the Japanese in B.C. are probably the most intelligent of all the racial groups which make up the total Canadian population" (p.509), Sandiford remarked that:

These facts have very great importance for B.C. Here we have an alien group whose mental capacity is greatly superior to that of the native stock. (p.508)

Social class differences. The discussion of intelligence was not confined to the expression of nativistic sentiments. The discussion also evinced the social class distinctions of the time:

There is a distinctive change of intelligence level with parental occupational grouping.... In all cases, children of professional workers head the list by a wide margin. Intelligence sufficiently high to achieve success in a profession is handed down to children. This is a matter of social significance. (pp.455-456)

The relationship between socio-economic status and intelligence reported by Sandiford in Putman-Weir simply mirrored results reported earlier for German, Belgian, and English children and continued to be a consistent finding in the literature. In a study of children in Regina during the 1960s, Kennett (1972) found evidence of the positive relationship between socio-economic status and intelligence test scores, favouring upper socio-economic status children. A similar relationship was obtained by Holmes

(1981) in a more recent study of children in British Columbia.

Eugenics. There is little doubt that eugenics provided part of the frame of reference for the interpretation of Sandiford's data. In the conclusion to the subsection concerned with socio-economic differences and intelligence, Sandiford expressed the connection clearly:

...the children of professional people are, on the whole, more intelligent than others and, eugenically, this is the group in which large families should be encouraged. (Putman and Weir, 1925, p.458)

Sandiford's notions about eugenics seemed benign in comparison to what was taking place in Alberta at approximately the same time. Describing the early eugenics movement in Western Canada, Chapman (1977) wrote:

In hopes of reforming society, Social Gospellers, Progressives and women's suffragettes advocated a philosophy of progress based upon the application of science to society which in this instance was eugenics. As support for such a theory permeated all levels of society, the post-1916 era saw the United Farm Women of Alberta adopt a eugenics programme in 1922. Six years later in 1928, the United Farmers of Alberta enacted Canada's first legislation concerning sterilization of the mental defective. (p.9)

Testing continues to figure prominently in Canada's education system. As of March 1985, the provinces of British Columbia, Alberta, Saskatchewan, Quebec, and Newfoundland administered compulsory final examinations to students at the secondary level. The number of examinable courses ranged from a low of seven in Alberta to nearly 150 in Quebec. In the five provinces administering compulsory final examinations, examination marks are combined with the marks awarded by teachers to determine the final marks awarded to students (Schulz, 1985).

Seven provinces -- British Columbia, Alberta, Manitoba, Quebec, Nova Scotia, Prince Edward Island, and Newfoundland -- are engaged in student assessment programs designed to assist in the evaluation of curricula and programs of study. Three of the seven -- Nova Scotia, Prince Edward Island, and Newfoundland -- use commercially prepared tests (Schulz, 1985). In most provinces, commercially published tests, including the Canadian Test of Basic Skills, Canadian Cognitive Abilities Test, Differential Aptitude Test, Gates-MacGinitie Reading Tests, Metropolitan Achievement Test, Stanford Achievement Test, and the Stanford Test of Academic Skills are used to make a variety of educational decisions ranging from those about the placement of individuals to the efficacy of programs (Educational Research

Institute of B.C., 1984).

The impact of standardized testing

Advertising copywriters have long known that the social imagery of a few people standing out from the crowd is a successful inducement to consume. The voice-over on a recent Chevrolet Monte-Carlo television commercial said it simply: "Put a little distance between yourself and the crowd." Although it has created enormous profits for advertisers and merchants who have used it, this social imagery was not created by a Madison Avenue genius. It finds its origins in the economic and social transformation that accompanied industrialization and in the rapid development of a technology for testing human abilities. Prior to industrialization one's position in society was affected primarily by birth rather than ability. Emphasis upon achievement and the development of a technology for measuring intelligence and other abilities provided a social environment in which three basic ideas could develop. These ideas shaped and continue to shape our thinking about the worth of human beings in our society. The three ideas are:

1. People should be judged by their abilities;
2. Abilities can be represented by a mathematical distribution which depicts a few people standing out from the mass of people; and
3. Human beings of extraordinary ability are not bound by the same conventions which bind "normal" people. (Sennett and Cobb, 1973)

The social imagery of a small group of people standing out from the crowd has had a profound affect on our view of education. The pattern of attitudes which grows out of this social imagery creates social, educational, and occupational consequences which are deeply disturbing.

Structure of tests. Standardized tests have traditionally served three major functions: evaluation, selection, and restriction. The growth of the testing movement has seen the development of a technology and a vocabulary that in some ways mystifies what the tests reveal about the people they are purporting to describe. The idea that the distribution of certain abilities in a population may be represented by a bell-shaped curve or "normal distribution" is not well understood. A bell-shaped distribution occurs in mathematics when the measurement of an entity is determined by the influence of many other factors, each of which are of equal and independent influence.

In the construction of standardized ability or intelligence

tests, developers seek to have the individual items which make up the test differentiate among the individuals taking the test. Test items are produced so that on some items only half of the people in a particular age group will answer the items accurately. The same population will answer correctly most of the items developed for younger people and will not answer correctly very many of the items developed for older people. In addition, the test items are developed to minimize the correlation among test items. As a consequence, the size of any given test score is determined by many independent responses of equal influence. This fact is obscured by the tendency to think about intelligence as if it were a measurable entity to which numbers can be assigned on a single value scale. Although we tend to think of intelligence as something tangible, it is, in fact, simply a convenient label for the sum of a person's responses to the individual items which make up the test.

Assumptions of test-makers. In approaching the construction of tests, developers assume that all people of a given age have had the same opportunity to develop the capacities being measured. This assumption is buttressed by the related assumption that the abilities being measured are genetically determined. These assumptions are bolstered by the development of tests which are **reliable**. Tests are developed in such a way that any given score does not change its position in relation to other scores over time. In other words, standardized tests are constructed to be insensitive to change (Feuerstein, 1979).

The construction of standardized tests of ability and intelligence proceeds on the basis of assumptions which run counter to the purposes of education. The process of education assumes that people change over time and that the rate and nature of the changes can be affected by systematic instruction. Standardized intelligence tests assume that what people are, today, is what they will be in the future. In other words, most standardized tests assume the inequities among people which they reveal are immutable and will continue to characterize the same individuals in the future.

The technology of standardized testing is of limited utility in improving student performance. It is nevertheless the case that standardized tests of intelligence and ability provide the basis for the comparison of student performances. And, such comparisons have inevitably led to selection. Based upon the notion that one should fit education to the needs of individual students, schools have used standardized tests of ability and intelligence as the means for deciding whether and what types of instruction should be made available to students. Standardized tests have been used as the basis for grouping elementary students within classrooms; for deciding whether secondary students should be allowed to pursue vocational, academic, or commercial programs; and for moving students out of the school system into

the work force.

The use of standardized tests for grouping elementary students within classrooms and streaming them into particular programs at the secondary level is based upon two related assumptions. The first is that there is a connection between student abilities and instructional method or program. The second assumption is that student achievement can be maximized if students of different abilities receive different types of instruction. Grouping by "intelligence" and ability has a long record of failure in producing increases in student achievement. Results of studies indicate quite clearly the groups of mixed ability do about as well as groups of students with similar abilities.

The notion of ability grouping is appealing: logic seems to dictate that teachers can teach most effectively when differences in student ability are minimal. However, several authoritative reviews suggest that ability grouping *per se* is unlikely to have a positive effect on student achievement. In fact, the typical results are that such grouping has no effect or only a minor positive effect on the achievement of students in the high-ability classes, but has clearly negative effects on the achievement and attitudes of students assigned to low ability classes. (Good & Brophy, 1984, p.273-274)

The opportunities of students from working class backgrounds in Britain, the United States, and Canada have been restricted by the use of standardized tests which tend to reinforce the disadvantages imposed by their economic background. By confirming the limitations imposed by background conditions, standardized tests make it more difficult for children from working class origins to reverse their economic situations through access to educational opportunities. Because people's lives are affected by the educational opportunities that are made available to (or withheld from) them, the decision to provide or withhold certain opportunities on the basis of standardized test scores becomes a political and economic decision.

The British "Eleven Plus" examination provides a clear example of the way tests have been used to determine which type of education the students would receive. Following the Second World War a tripartite system of secondary education was established and a battery of tests was administered to all students in publicly financed education. Students who performed well on the examination went on to the grammar schools which offered an academic curriculum leading to a professional occupation. Students who performed less well were assigned to technical schools which offered a curriculum leading to skilled technical occupations. The students who performed least well, by far the

largest group of students, were assigned to secondary modern schools which offered a general curriculum that was designed to prepare them for the unskilled jobs they were expected to fill.

Conclusions

During most of this century, examinations were used in such a way as to control access to the secondary and post-secondary education for the majority of students in Canada. The demands of working class and immigrant groups for better and more education were blunted by the use of standardized tests. The "scientific" status of these tests provided justification for the restriction of educational opportunities. Today, the practice of making opportunities available on the basis of what are typically regarded as in-born abilities is as strong as it was a half-century ago.

All students should have access, within the constraints of what a given school can offer, to a program of studies that is challenging, **yet appropriate to their needs, abilities, and aspirations.** (B.C. Ministry of Education, 1984, p.10)

Today, members of particular ethnic and economic groups are consistently over-represented in some programs and under-represented in others. In Canada, participation rates in post-secondary education differ substantially by social class and region, revealing an inequitable distribution of opportunity for young people from different backgrounds (Guppy, 1984).

When the basis for educational decisions rests upon information provided by standardized tests of intelligence and ability, we can be certain that the inferences being drawn from the test data are influenced by some estimation of the person within an economic, political, or social frame of reference. Because of their bias in favour of white, middle-class students and their use in determining the opportunities available to students in school and people in the workforce, standardized tests have helped to perpetuate discrimination against people from low income and minority backgrounds and to preserve existing social inequalities.

REFERENCES

- Chapman, T.L. (1977). Early eugenics movement in Western Canada. *Alberta History*, 25, 9.
- Collini, S. (1979). *Liberalism and sociology*. Cambridge: Cambridge University Press.
- Feuerstein, R. (1979). *The dynamic assessment of retarded performers*. Baltimore: University Park Press.
- Galton, F. (1869). *Hereditary genius: An inquiry into its laws and consequences*. London: Macmillan.
- Good, T.L. & J.E. Brophy. (1984) *Looking into classrooms (third ed.)*. New York: Harper & Row.
- Graduation '87. (1984). (Mimeographed). Victoria, BC: Ministry of Education.
- Guppy, N. (1984). *Education under siege: financing and accessibility in B.C. universities*. (Mimeographed). Vancouver, BC.
- Holmes, B. (1981). *Individually administered intelligence tests: An application of anchor test norming and equating procedures in British Columbia*. (Unpublished dissertation). Vancouver, BC: University of British Columbia, Faculty of Education.
- Hofstadter, R. (1969). *Social Darwinism in American thought*. New York: George Braziller.
- Kamin, L.J. (1974). *The science and politics of I.Q.* Toronto: John Wiley & Sons.
- Kennett, K.F. (1972, March). Intelligence and socioeconomic status in a Canadian sample. *The Alberta Journal of Educational Research*, XVIII, 1.
- Martindale, D. (1960). *The nature and types of sociological theory*. Cambridge, MA: The Riverside Press.
- Putman, J.H. & G.M. Weir. (1925). *Survey of the school system*. Victoria, BC: Charles F. Banfield, Printer to the King's Most Excellent Majesty.
- Schulz, H. (1985, March). Summary of provincial assessment practices in Canadian public education. Produced for the Council of Ministers of Education, Canada.
- Sennett, R. & J. Cobb. (1973). *The hidden injuries of class*. New York: Vintage Books.
- Spearman, C. (1904, April). "General intelligence", objectively determined and measured. *American Journal of Psychology*, XV, 2.
- Testing service reference guide. (1984, September). Vancouver, BC: Educational Research Institute of British Columbia.
- Terman, L.M. (1916). *Stanford revision of the Binet-Simon scale*. Boston: Houghton Mifflin.
- Terman, L.M. & M. Merrill. (1937). *Revised Stanford-Binet scales*. Boston: Houghton Mifflin.