Lessons from the history of intelligence testing

David F. Lohman
The University of Iowa
May 2, 2003

Unexpurgated version of a paper that appears in the International Journal of Educational Research, 27, 1-20.

What can we learn from the history of intelligence testing that might improve future assessments? Much. But how to tell it? Anyone who has read both old and new could tell tales of insights overlooked or repeatedly rediscovered, of long and carefully conducted research programs that were lost entirely or reduced to brief, often inaccurate caricatures in modern summaries, of roads not taken that would have changed the whole. I will not attempt such a litany here. Instead, for no better reason than that three seems to be the magic number in the literature on human intelligence, I have identified three general themes. I list them here, and then elaborate each in turn. The first and last points require relatively brief elaboration. The second point is more difficult to document, and so it comprises the bulk of this article.

The general themes are:

Theme 1. The developers of intelligence tests were not as narrow minded as they are often made out to be; and, as a necessary corollary, nor are we as clever as some would have it. Those who have not themselves read widely from the books and articles of luminaries such as Binet, Spearman, Thorndike, and Stern are not so much condemned to repeat history (as Santayana claimed) as they are to say and write silly things. Nevertheless, partly because of our ignorance of the past, but mainly because there are larger tides in the affairs of humankind, controversies about intelligence repeat themselves. Transitions that look new in the short run often look familiar in the long run. We may never step into the same river twice, but we do have a habit of rediscovering the same stepping stones and potholes therein.

Theme 2. Then and now, theories of intelligence are the product not only of data and argument, but also of the personal proclivities and professional experiences of theorists, of their beliefs about what science is and how it should be conducted, and of the larger social, political, and religious themes that form the fabric of the cultures in which they live.

Theme 3. Some of the most important changes in the format of intelligence tests were administered were dictated more by the demands for efficiency and reliability than by psychological theory. Particularly noteworthy were (a) the shift from individually administered tests to large batteries of group-administered, paper-and-pencil tests, and (b) the shift from tests that required the examiner to make judgments about how items or tasks were solved to tests in which examiners merely tabulated the number of items solved correctly. The underlying tension here was (and continues to be) between those who place higher priority on the statistical properties of test scores -- particularly their factor structure and reliability -- than on the psychology of the tasks.

Theme I. Study the classics

The life work of a great and productive scholar cannot be compressed into a few sentences. Those who take the time to read (or re-read) Spearman or Binet or Thorndike will be struck by how much larger their views were than even the best summary suggests. The surprise is greatest if one's reading has been limited to accounts in textbooks and other secondary sources. For example, I have been concerned for several years now with the effects of practice on mental tests, particularly tests of spatial abilities. With the notable exception of Ackerman's studies of individual differences in skill acquisition

(see, e.g., Ackerman, 1987), nowhere in that literature have I seen references to the extensive review of practice and transfer effects that Thorndike reports in his 1913 Educational Psychology text. If Thorndike is mentioned at all, it is in a fairly standard comment (often dismissive and usually inaccurate) about the theory of identical elements. Those who take time to read the original will find much more.

More importantly here, Thorndike's studies of practice and transfer were pivotal in his attempts to develop a better intelligence test than Binet had assembled. In the end, though, he used well-practiced school tasks in his intelligence tests because his studies showed large practice effects when students attempted novel problems. So called "performance tests" were particularly labile to practice. A generation of psychologists trained after the development of the Wechsler scales seems not to have worried much about such problems, even though practice effects average about 0.6 SD on the performance scales of the WISC-R (Cronbach, 1990, p. 277).

Nor do modern advocates of Galton's methods of measuring Spearman's 'g' seem to have read the extensive prologue to Spearman's (1904a) paper in which he discusses in considerable detail the effects of differential practice on such measures, such as Binet's studies with RT and other tasks, or Seashore's studies with pitch and loudness discrimination. Binet concluded that the relationship between such measures and intelligence is strongest on the first trial, and diminishes with practice. Many years later, Fleishman and Hempel (1954) and Ackerman (1988) offered elaborations of this theme. In a related vein, Carroll (1987) argues that the correlation between the variability of RT's and g in Jensen's (1982) studies reflects not noise in neural conductivity, but variations in attention. Binet (1903) offered the same hypothesis to explain why children's best RTs compare favorably with those of adults, even though their mean RT is much longer.

The important point, however, is not that hypotheses recently advanced have been advanced before. Rather, it is that the corpus of good writing and research by early developers of intelligence tests is not only much larger, but much more variegated than our simple summaries suggest. Binet was maddeningly eclectic, especially to those like Spearman who were trained to build coherent psychological theories: "It would seem as if, in thus inconstantly flitting hither and thither, Binet can nowhere find theoretical perch satisfactory for a moment even to himself" (Spearman, 1923, p. 10). Spearman also had wider views than the standard account of the two-factor theory of intelligence suggests. His classic 1904 paper "'General intelligence,' objectively determined and measured" alone runs 92 journal pages! And Thorndike was even more prolific. Even his grandson, Robert M. Thorndike, was surprised by the breadth of his grandfather's work on human intelligence after he had time to immerse himself in it (see Thorndike & Lohman, 1990, p.v).

Thus, it is not only our individual memories that distort, simplify, and forget. Our collective memory -- as captured in the brief summaries of past errors and accomplishments litanized in the introductions of countless articles and chapters -- reinforces key-hole views of a past that is much more complex and untidy than we would have it. More is at stake here than constructing a better historical legacy. As he completed his massive review of factor-analytic studies of human abilities, Carroll (1989) was struck by the non-cumulativeness of the enterprise. Important lessons learned at one time seemed to be lost on a new generation. Why? Perhaps we have been lulled into

believing in the metaphor of scientific progress. If the march is ever upward, then the value of research varies inversely with its recency. Another possibility is the increasing dominance of statistical methods over psychological investigation. But that is another part of the story to which I will return later. Here I can do no more than to encourage readers to read -- and if they have already read -- to re-read classic and obscure papers. If nothing else, the papers are generally easy -- even fun to read. Argument and data took precedence over statistics, which were often crude or absent entirely. Indeed, with a modern personal computer and a good statistics package, one can profitably rework old data, as Carroll (1993) has shown us.

Finally, although brilliant and creative and sometimes maddeningly inconsistent, the developers of intelligence tests were firmly rooted in the academic, social, and political climate of their time. Because of this, and because of the enormous energy most expended on their psychological research, non-specialists who took a broader view of society often saw the dangers and hidden assumptions of tests more clearly than did their developers.

> [I am] impressed by the discovery that more than once [in the history of intelligence testing] nonspecialists, like Walter Lippman . . . seem to have had a better grasp of the real issues involved, in spite of misunderstandings of technical details, than the scientists themselves. War may be too important to be left to the generals. (Samelson, 1979, p. 141).

Which brings us to the next theme.

Theme II. The role of personal beliefs

Scientists as Believers

There are those who believe that the history of science -- particularly of the social sciences -- cannot be understood without some insight into the social and psychological factors that affect scientists.[1] For example, in his controversial critique of intelligence testing, Gould (1981) claims:

> My message is not that biological determinists were bad scientists or even that they were always wrong. Rather, I believe that science must be understood as a social phenomenon, a gutsy human enterprise, not the work of robots programmed to collect pure information.[2] (p. 21)

In the same way, but with a less muckraking tone, Fancher (1985) claims that the story of intelligence testing comes into focus only when one begins to understand the impact of personal experience and belief on the theories publicly advocated. He begins

---

[1] This is but one of five models that Richards (1981) discusses. The other four describe science as the product of (1) genius, inspiration, or even revelation, (2) a slow but steady march down the road of progress, (3) Kuhnian paradigm shifts, and (4) an epistemological Darwinism of sorts, in which ideas are generated in response to expectations not met, but are also influenced by the scientist's culture, context, and psychological disposition.

[2] James Watson (1980) makes a similar point in his account of the discovery of the double helix: "Science seldom proceeds in the straightforward manner imagined by outsiders. Instead, its steps forward (and sometimes backward) are often very human events in which personalities and cultural traditions play major roles." p. xi

with the disparate childhood experiences of John Stuart Mill, who was educated by his father and shielded from knowledge of his precocity, and Francis Galton, who early in life learned that his achievements were unusual and who in his public education constantly compared himself to others. Mill later championed the environmentalist perspective and argued that one should resort to biological explanations of individual and group differences only after all reasonable environmental explanations had been explored and refuted. Galton, on the other hand, saw individual differences as largely genetic in origin and eugenics as the path to improvement. These two perspectives on the origin of individual differences in intelligence can be seen to one extent or another in the personal beliefs of many of the protagonists in the story of intelligence testing, such as Alfred Binet and Charles Spearman.

Both Binet and Spearman came to psychology after they had finished their university work. Spearman was an officer in the Royal Engineers; Binet a lawyer who had never practiced law and who had dropped out of medical school. Spearman later called the decision to join the army "the mistake of my life" (Spearman, 1930, p. 300). Binet was even more sarcastic about his early decision to study law, calling it "the career of men who have not yet chosen a vocation" (quoted in Wolf, 1973, p. 3). Yet their reactions to Mill's associationistic psychology were quite opposite. Binet called Mill his only master in psychology (Wolf, 1973). Not so for Spearman. He would later recall that "my initial reaction to [the argument of Mill and other associationists that experience was the foundation of all knowledge] was intensely negative... My conviction was accompanied by an emotional heat which cannot ... be explained on purely intellectual grounds." (Spearman, 1930, p. 301). Spearman's reaction is not uncommon, although his report is. Scientists routinely harbor the belief that somehow their views should be justified on a purely rational basis. Affect is considered a corrupter of cognition; theorists should strive to rise above such distractors.

Damasio (1994) takes a different view. His investigations of patients with damage to the ventromedial sector of the prefrontal cortex show the positive contributions of affect to rational cognition, particularly the ability to solve ill-structured problems.

> Reason may not be as pure as most of us think it is or wish it were . . . More directly, emotions and feelings may not be intruders in the bastion of reason at all; they may be enmeshed in its networks, for worse and for better..... In fact, the absence of emotion and feeling is no less damaging, no less capable of compromising the rationality that makes us distinctively human and allows us to decide in consonance with a sense of personal future, social convention and moral principle. (p. ii - xii)

This is not a claim that rational processes are not important. Feelings do not solve a problem.

> At their best, feelings point us in the proper direction, take us to the appropriate place in a decision-making space, where we may put the instruments of logic to good use. We are faced by uncertainty when we have to make a moral judgment, decide on the course of a personal relationship, choose some means to prevent our being penniless in old age, or plan for the life that lies ahead. (p. xiii)

A good theory invariably reaches beyond the information given; through it the theorist attempts to impose a new order on what frequently is an ill-structured problem. At the very least, then, the creation (and, for later readers, acceptance) of a theory is influenced by affect. In the extreme, the solution may merely provide rational justification for an emotional reaction to an event. More commonly, affective reactions and the beliefs they entail color the way we interpret ambiguous data.

Mill, Galton, Binet, and Spearman are by no means the only examples of this phenomenon. Although exceptions are noteworthy, the rule is that a theorist's earliest pronouncement about the relative importance of nature or nurture in intelligence differs little if at all from the one made at the end of a career. For example, in his first published article "Experimental tests of general intelligence" Burt (1909) concluded that because the thirteen upperclass boys in his study outperformed the thirty lowerclass boys on tests he thought unaffected by practice, intelligence must be inherited "to a degree which few psychologists have hitherto legitimately ventured to maintain" (p. 176). By 1911, Burt had defined intelligence as "allround innate mental efficiency," (quoted in Hernshaw, 1979, p. 49), a view to which he adhered throughout his career. "It was for [Burt] almost an article of faith, which he was prepared to defend against all opposition, rather than a tentative hypothesis to be refuted, if possible, by empirical tests." (Hernshaw, 1979, p. 49).

Terman (1906) showed his hand in his dissertation "Genius and stupidity: A study of some of the intellectual processes of seven 'bright' and seven 'stupid' boys." Near the end of the dissertation, he speculated: "While offering little positive data on the subject, the study has strengthened my impression of the relatively greater importance of endowment over training, as a determinant of an individual's intellectual rank among his fellows." (p. 372, italics original) Once again, experience seemed not to alter these early beliefs, as later clashes with Lippman and the Iowa group showed.

Exceptions are noteworthy. Brigham (1930) publicly retracted his early, hereditarian interpretation of ethnic differences in intelligence based on the U.S. Army data from WWI. However, the retraction came because the data would not support the conclusions rather than because the conclusions themselves had changed (Cronbach, 1975). J.M. Hunt recounts an experience that is a better example. His book Intelligence and Experience (1961) summarizes research on the effects of experience on the development of intelligence. (At the time, it also served to introduce the work of Piaget to many U.S. psychologists.) In trying to explain why belief in a fixed intelligence was so prevalent, Hunt appealed to Festinger's (1957) theory of cognitive dissonance:

> In his own professional life history, the writer finds in himself some evidence of [cognitive dissonance]. So long as he was professionally identified with the testing function, it was highly comforting to believe that the characteristics tested were fixed in individuals. Evidence hinting that these characteristics were not fixed produced intense dissonance, for it threatened his belief in fixity and the adequacy of his professional function as well. Such a factor may help to explain the sometimes excessive violence of the polemics concerning the constancy of the IQ and the effects of training that were common in the years just (prior) to World War II. (pp. 14-15)

Surely cognitive dissonance may help explain why it is difficult to change beliefs. But it does not explain the "violence of the polemics." That seems better to reflect the fact that beliefs are often grounded in affect rather than merely tinged with it.

Political and Social Climate

Psychologists (including this one!) are not historians. The tales we tell each other about the origins and development of mental testing are often remarkable for their failure to consider the influence of larger political and social influences on testing. With increasing regularity, we acknowledge the impact of the broader culture on cognition and cognitive development. For example, Bronfrenbrenner (1979) argues that abilities develop through a child's interactions not only with her immediate social environment but also with the attitudes and ideologies of the broader culture in which she lives. Cultural relativism, long relegated to the extreme left-wing of psychometrics, is now afforded a respected place at the table (e.g., Irvine & Berry, 1988; Laboratory of Comparative Human Cognition, 1982) and a central role in at least one major theory of intelligence (e.g., Sternberg, 1985). Yet the stories we tell about the development of theories within the discipline are often remarkably devoid of such influences. In the preface of their account of intelligence testing in Britain, Evans (a psychologist) and Waites (a historian) note:

> Most histories of psychology have been written by professional psychologists with a strong commitment to the body of belief accepted as achieved knowledge within their profession ... Such histories are not necessarily uncritical, but the critical point of view is very restricted. Past error appears merely as a series of hurdles successfully overcome on the road to current theory. (1981. pvii)

Modern conceptions of intelligence were birthed in the second half of the 19th century. It is therefore impossible to understand either the theories that describe the construct or the tests developed to measure it without some understanding of the political and social ideology of the time. However, identifying the starting point is much like stepping into a stream at a particular point--perhaps where it rounds a bend--and declaring "We will take this as the beginning." Such fictions clearly mislead, but one must begin somewhere. Herbert Spencer provides such a convenient starting point.

Social Darwinism. Spencer advocated a theory of evolution before Darwin (1859) published his Origin of Species. Spencer's argument, based as it was on philosophical, anthropological, and geological speculation, was largely ignored; Darwin's biologically-based argument was not. Spencer soon allied himself with Darwin's theory and sought to apply the theory to the full range of human knowledge and endeavor. The effort resulted in the ten volumes of his Synthetic Philosophy (which included The Principles of Psychology).

Spencer saw evolution as the key to all science. Evolution, he said, proceeds from incoherent homogeneity, as is found in the lowly protozoa, to coherent heterogeneity, as is found in mankind and the higher animals.[3] More importantly,

[3] William James, the once-believer-turned-critic of Spencer (who used Spencer's text in his psychology classes before writing his own) parodied this view to his classes as "Evolution is a change from a no-howish untalkaboutable all-alikeness to a somehowish and in general talkaboutable not-all-alikeness by continuous stick-togetherations and somethingelseifications." (cited in Hofstadter, 1983, p. 129)

heterogeneity increased unidimensionally. In this both Spencer and Darwin followed the lead of Locke and Leibniz in viewing all life as falling along a continuous, unbroken scale. Leibniz said it most succinctly: Natura non facit saltum ("Nature does not make jumps"). For Spencer, the tick marks on this scale marked increases in intelligence: "Gradually differentiated from the lower order of changes constituting bodily life, this higher order of changes constituting mental life assumes a decidedly serial arrangement in proportion as intelligence advances" (1897, p. 406).

The idea of a serial order--not only between species but within humankind--was brought about by an unfettered competition among individuals. "Survival of the fittest" is his phrase, not Darwin's. If species evolved by becoming more intelligent, then to help humankind become more intelligent was to further the work of evolution. Yet Spencer held firmly to a Lamarckian view of the heritability of acquired characteristics. His theory was thus optimistic about the value of education. If parents transmitted to their children through their genes some remnant of their own education, then problems of social degenercy could be solved in a few generations. Those who dismissed Lamark, however, saw the answer in eugenics rather than in education.

Galton was the chief exponent of this more structured view of human inequality. Like other liberals of his day, Galton advocated a meritocracy in which income and position within the social hierarchy would be based on innate ability rather than on parental social status. In Hereditary Genius (1869), he argued that mental characteristics were inherited in the same manner as physical characteristics. What he lacked, though, was a way to measure innate ability so that individuals could be properly placed in the hierarchy. Tests could not fulfill this function unless they truly measured innate ability, and on a single dimension. Those convinced of the need for eugenics were thus more willing than they otherwise might have been to believe assertions that the new intelligence tests (which, parenthetically, confirmed their natural superiority) measured innate, general intelligence.

Advancing the cause of the able was only a part, and for many, a lesser part of the problem. The specter that most haunted American and European intellectuals during this period was the prospect the degeneration of human kind and subsequent collapse of society. "Indeed, 'degeneration' was arguably the most potent concept in the medical and biological sciences of the period" (Wooldridge, 1994, p. 20). In Britain, politicians and military leaders blamed defeat in the Boer War on the physical unfitness of the masses. In Italy, Lambroso warned about the return of the primitive, especially "natural criminals" (cf. more recent discussions of a permanent "underclass" in the U.S., as in Hernstein & Murray, 1994). In Germany, fear of racial degeneration and the quest for racial purity found their way into medicine, the social sciences, and political thought. In France, the medical/psychiatric concept of dégénerescence pervaded political debates about national decline.

> From Jaurès to Maurras, political discourse was obsessed by the question of national defeat and the ensuing chaos; . . . alcoholism, habitual crime, and depravity were cast through the image of a social organism whose capacity for regeneration was in question. National defeat, degeneration and social pathology appeared to be caught up in an endless reciprocal exchange. (Pick, 1989, p. 98)

A declining birthrate in France and stagnant birthrates in Germany and Great Britain seemed to confirm fears that degeneration had set in. Governments took note because military hegemony was in peril for a nation that could not raise armies larger than its rivals, especially from a population of "stunted, anemic, demoralized slum dwellers. " (Wooldridge, p. 22)

The novels of Zola and the plays of Eugène Brieux, whose Damaged Goods inspired English-speaking authors such as Bernard Shaw and Upton Sinclair, kept these issues squarely before the public. The proximal causes of degeneration were (a) the higher birth rate among the lower classes and "races", especially the teeming masses of urban poor, and (b) the movement of "races" outside of their natural ecologies. Thus, an African

> placed outside of his 'proper' place in nature--too stimulating an intellectual or social environment, or in a climate unsuited to his 'tropical' nature--could undergo a further 'degeneration,' causing the appearance of atavistic or evolutionarily even more primitive behaviors and physical structures. (Stephan, 1985, p. 98)

Similarly, those of the white race who lived in warmer climates risked becoming diseased and anemic, sexually promiscuous, and culturally backward. Colonizers needed to make all efforts to maintain the home environment, to avoid unnecessary contact with the native populations, and to return home whenever possible to restore their "type" and repair degeneracy acquired abroad.

In America, emancipation of the slaves gave new urgency to the question of the boundaries between whites and black. Blacks who attempted to move away from the warmer southern states or to advance beyond their “natural condition” of servitude would degenerate and eventually die out. The statistician and economist Hoffman argued that blacks were "a race on the road to extinction," and thus whites--who thrived in the political and geographical climate of America--need have no fears of a newly freed black race (see Stephan, 1985).

The racial mixing also resulted in degeneration. If nature separated humans into ranks, then mixing them was not only unnatural, but an invitation to atavism. "Unnatural unions" produced a hybrid that was inferior to either parent. In America, and later in Germany, the doctrine of rassenhygiene (race hygiene) became the watchword for a group that looked back to a braver, purer Teutonic past as much as it looked forward to a eugenically improved race.

Thus, the specter of a complete collapse of society haunted many American and European intellectuals during this period. Evidence of the social, moral, and physical decay of humanity seemed irrefutable. In the U.S., many feared that the massive immigration of Europe's poor and degenerate was nothing short of complete folly. Nevertheless, the American and European agendas differed. Whereas European intellectuals (particularly in England) continued to struggle against inherited privilege. Americans in- and outside of academia were more concerned with the social upheavals threatened by immigration and the explosive growth of mass education. The first led to ready acceptance for eugenic proposals, and the second for more efficient ways to conduct the business of education. Both of these agendas conflicted with the principle of equality etched in Jefferson's Declaration of Independence, and thus set the stage for later debate about whether tests that showed differences between ethnic and social classes

were, ipso facto, biased. Even more important, Americans then and now seemed less uniformly committed to the notion of a single rank order as their British cousins.[4] Thus, E. L. Thorndike found no conflict in on the one hand, advocating a genetic basis for human intelligence while, on the other hand, arguing for the existence of several intellectual abilties rather than a single, general factor.

Educational reforms. Intelligence testing became a part of the British educational system because it advanced a meritocratic agenda; it became part of the American system because it helped solve practical problems in an educational system overrun with pupils. Schooling was expanding exponentially. From 1890 to 1918 the population of the U.S. increased 68%. High school attendance during this period increased 711%. On average, more than one new high school was built every day during this period (Tyack, 1974, p. 183). However, school curricula still defined a Procrustean bed in which "the wits of the slow student were unduly stretched and. . . of the quick pupils amputated" (Tyack, p. 202). Many started, but few finished. Ayres (1909) showed that the number of students in each grade dropped precipitously between first and eighth grade. "The general tendency of American cities is to carry all of their children through the fifth grade, to take one half of them to the eighth grade, and one in ten through high school." (Ayres, 1909, p. 4). And of those who remained in school, many in the early grades were "laggards" who had been held back. The culprit was thought to be the failure of the system to adapt itself to the intellectual abilities of its students. A variety of methods for classifying students or adapting the pace of instruction had been used in American schools for many years (Chapman, 1988). But the intelligence test was heralded as a more scientific and efficient method of performing the task. Tests then as now provided a means for sorting and classifying people that was ostensibly objective and fair. And what could be fairer than an educational system that was adapted to the natural ability levels of its students? Intelligence tests helped administrators in a school system newly infatuated with a corporate model of centralization, bureaucratization, and efficiency, perform and defend the sorting functions they were asked to perform. Test publishers thus found a ready market for their products. Terman estimated that probably a million children were given a group intelligence test in 1919-1920, and two million the next. Even Walter Lippmann, who is now remembered by psychologists chiefly for his debates with Terman about intelligence testing, applauded the goal:

> The intelligence test promises to be more successful in grading the children. This means that the tendency of the tests in the average is to give a fairly correct sample of the child's capacity to do school work. In a wholesale system of education, such as we have in our public schools, the intelligence test is likely to become a useful device for fitting the child into the school. This is, of course, better than not fitting the child into the school, and under a more correct system of grading, such as the intelligence test promises to furnish, it should become possible even where education is conducted in large classrooms to specialize the teaching, because the classes will be composed of pupils whose capacity for school work is fairly homogeneous. (Lippmann, 1922, in Block & Dworkin, 1976, p. 18)

---

[4] French observers of 19th century English society hyothesized that, from birth, the English seemed afflicted with la mentalité hiérarchique (Tawney, 1952, p. 23).

Nativism. The practical use of tests, however, has always followed a path largely independent of theoretical debates about the nature of intelligence. In America (as in England) social darwinism was the dominant view. In fact, "Spencer's writings were so dominant in discussions of society and politics that they virtually sank into the unconscious of American political deliberation, ceased to be an argument, became obvious, and became common sense" (White, 1977, p. 36-37).[5] Mass immigration of peoples from southern and eastern Europe (on the Atlantic coast), and from China, Japan, and the Philippines (on the Pacific coast) led American intellectuals to emphasize the racial or ethnic differences in ability more than Europeans, who were more concerned with social stratification within their societies. The new immigrants were, on the East Coast, poorer and more Catholic than before. More important, they did not as readily assimilate into the existing order, but instead insisted on keeping their own language and customs. Catholics set up their own schools, and achieved political control in some cities of the Northeast, thereby reviving Protestant fears of Papal influence. Nativist political organizations fought against these trends.

> Previously vague and romantic notions of Anglo-Saxon peoplehood were combined with general ethnocentrism, rudimentary wisps of genetics, selected tidbits of evolutionary theory, and naive assumptions from an early and crude imported anthropology (later, other social sciences at a similar stage of scientific development
>
> added their contributions) to produce the doctrine that the English, Germans, and others of the 'old immigration' constituted a superior race of . . . 'Nordics' or 'Aryans.' (Gordon, 1964, p. 97)

These beliefs were given wide currency in Madison Grant's (1916) The Passing of the Great Race. The obvious implication of such views was to exclude the inferior races, an end finally achieved in the restrictive immigration laws of the 1920's, but only after nativist sentiments were augmented by a post-war isolationism.

But what to do about those already admitted? Ellwood Cubberly, Terman's dean at Stanford, echoed the feeling of many:

> These southern and eastern Europeans are of a very different type from the north Europeans who preceded them. Illiterate, docile, lacking in self-reliance and initiative, and not possessing the Anglo-Teutonic conceptions of law, order, and government, their coming has served to dilute tremendously our national stock, and to corrupt our civic life. The great bulk of these people have settled in the cities of the North Atlantic and North Central states, and the problems of proper housing and living, moral and sanitary conditions, honest and decent government, and proper education have everywhere been made more difficult by their presence. Everywhere these people tend to settle in groups and settlements, and to set up here their national manners, customs, and observances. Our task is to break up these groups or settlements, to assimilate and amalgamate these people as a part of our American race, and to implant in their children, so far as can be done, the Anglo-Saxon conception of righteousness, law and order, and popular

[5] Wooldridge (1994, p. 20) quotes Victor Berard as claiming that social darwinism had likewise "entered into the very bones of the [British] nation."

> government, and to awaken in them a reverence for our democratic institutions and for those things in our national life which we as a people hold to be of abiding worth. (Cubberly, 1909, pp. 15-16)[6]

Eugenic proposals. If the survival of a democracy depends on the ability of its citizens to make intelligent decisions, and if intelligence is innate, and individuals and groups differ in intellectual competence, then it was the moral duty of those who would improve (or at least not openly contribute to the degeneration of) humankind to restrict immigration of such peoples into the country and to stem their proliferation within society. Alarmist (and later, sober) reports of the dysgenic effects of the higher birthrate among the poor and the less intelligent led to calls for sterilization of the retarded and restrictions on immigration (see, e.g., Cattell, 1940). As late as 1940 in his massive Human Nature and the Social Order, E.L. Thorndike was writing that although

> much in the legislation concerning sterilization is doubtless misguided, . . . on the whole its operations seem more beneficent than those of an equal amount of time and skill spent in 'social education.' Indeed the first lesson in social education for an habitual criminal or a moral degenerate might well be to teach him to submit voluntarily to an operation that would leave his sex life unaltered but eliminate his genes from the world. The same would hold for dull or viscous epileptics and for certain sorts of dull and viscous sex perverts. (p. 455)

Although psychologists certainly contributed to the discussion of eugenic proposals, they were not the only or even the most important voices. For example, contrary to popular opinion, neither psychologists nor their army testing data exerted much influence on the restrictive U.S. Immigration Law of 1924.

> [The immigration law of 1924] was the culmination of efforts begun in the 1890s and supported by a far-flung coalition of forces from the Immigration Restriction League all the way to the American Federation of Labor. Much of the power of this movement was based on economic issues. . . . It was predominantly the biological argument of the eugenicists and racists under the leadership of Madison Grant and C. B. Davenport that produced the scientific . . . legitimation [for this legislation]. (Samelson, 1979, pp. 135-136).

Eugenics was part of a larger Zeitgeist that had at its core a belief in the improvement of humankind through the application of the scientific method to the study of people and their institutions. Interventions ranged from the child study movement and the enactment of child labor laws, to the application of corporate methods to education, to time-and-motion studies of industrial production, to the enactment of eugenic proposals into sterilization laws. It was at root, though, a reaction to a widespread conviction that evolution was going in reverse, that all that was good and noble and worth saving was slowly, inexorably sinking back into violence and degradation of an earlier level in the evolution of humankind. It was more reactionary than visionary, more backward than forward looking, more a product of fear than of hope.

---

[6] The trail does not stop here. Stanford's President David Starr Jordan was a well-known biologist and leader in the eugenics movement.

The religious context. Finally, the ground had been prepared for the seed of intelligence testing by an even larger and earlier cultural movement: the Reformation. Salvation, Luther said, was not to be achieved through good works but through grace. Those thus saved were the new chosen people, an analogy taken quite literally in Calvinist Holland (Schama, 1987). If some were elected for salvation, then others were predestined for damnation--at least in Calvinist and Puritanical writings.[7] It is now acknowledged that these beliefs influenced an astonishing array of other--often distant--aspects of the social, political, and economic activity of these peoples and the cultures they influenced. For example, Weber (1904/1958) argued that, paradoxically, the belief in predestination fueled the economic enterprise of ascetic Protestant sects (such as the Puritans). Schama (1987) claims that the sense of self-legitimation as a "chosen" people of the sort that pervaded Dutch culture during the seventeenth and eighteenth centuries also "helps account for the nationalist intransigence of . . . the Boer trekkers of the South African Veldt, the godly settlers of the early American frontier, even the agrarian pioneers of Zionist Palestine" (p. 35). If so, then in America, nativism and manifest destiny were not far behind. More important, the belief that some are predestined for salvation and others for damnation is not only compatible with the belief that only some are chosen intellectually, but more significantly that such gifts might properly be used as an arbiter of individual merit and worth. "We are comfortable with the idea that some things are better than others," proclaim Herrnstein and Murray (1994, p. 534). But they are also comfortable with the fact that some people are better than others, that the best measure of better is IQ, and that a meritocracy based on intelligence is "what America is all about" (p. 512). Perhaps it is, and perhaps that is why the study of individual differences in general and of intelligence in particular has been more popular a topic in countries where Calvinism and Puritanism once flourished than in countires where such beliefs never attained a significant following.[8]

The meritocracy. In education (Tyack, 1974), mental testing (Cronbach, 1975), and politics (Gardner, 1961), it is one of the strange ironies of history that reformers

---

[7] It is useful to distinguish between the simple predestination of Paul (some are predestined for salvation) from the double predestination of Augustine, Luther, and Calvin (some are predestined for salvation, others for damnation). The latter view is more congenial with eugenic proposals to eliminate the unfit. However, Luther's emphasis on the equality of all men before God had a more profound impact on the societies that followed his lead rather than Calvin's.

[8] The medieval church certainly recognized differences among people in social status, but its view differed from the post-Darwinian view in two major respects: First, from Aristotle through Aquinas it was believed that humans differed qualitatively from other animals by virtue of their reason, intellect, or immortal souls. The classic definition of man as a rational animal saw the possession of a rational facully as the key difference not only between man and other animals, but among men. As one defender of this orthodoxy put it "Once we step down from that tall and splintered peak of pure insanity we step on to a tableland where one man is not so widely different from another. Outside the exception, what we find is the average" (Chesterton, 1922, p. 38). Second, differences among men in learning or wealth or social status--while important in this world--had no bearing on their status in the next. The Anglican Jeremy Taylor put it thus:

> If a man be exalted by reason of any excellence in his soul, he may please to remember that all souls are equal, and their differing operations are because their instrument is in better tune, their body is more healthful or better tempered; which is no more praise to him than it is that he was born in Italy (in Tawney, 1954, pp. 37-38).

often misjudge the consequences of their reforms. Those who sought the abolition of a society artificially stratified by hereditary privilege with one based on merit did not foresee that a meritocracy would create new problems, or that judgments of merit would lead to new inequalities. Who would have guessed that the tests which were heralded in one generation for opening the doors of higher education to all would, in the next, be attacked by some as artificial gatekeepers?

Tests do predict outcomes, and they are surely a fairer way of allocating scarce resources than any other alternative yet devised. But they see through the glass only dimly. Popular and professional misconceptions that tests measure (or presently might measure) ability uncontaminated by culture or social class or motivation continue to plague interpretations of test scores, as the recent debate on affirmative action in the U.S. shows. Few understand complexities such as the social class bias in schooling itself (Davis, 1949), the verbal-educational bias in many criterion measures of job performance, particularly test scores or supervisor ratings gathered immediately after training (Frederiksen, 1984), or how the generally weak relationships between predictor and criterion in personnel selection mitigates arguments against proportional within-group hiring (Cronbach & Schaeffer, 1981). Selecting those most likely to succeed in the current system helps perpetuate that system. This may be good, but it is no small irony that the same tests which help liberate talent also help conserve the system. Renewed arguments for the use of measures of general ability in personnel selection (Schmidt, Hunter, & Pearlman, 1981) are particularly important since a "single-rank-order selection is only a shade less conservative than the aristocratic selection it replaced, since to a significant degree it also perpetuates advantage of birth" (Cronbach & Snow, 1977, p. 8).

The yardstick by which merit is measured reflects the needs and demands of society at a particular point in time. If exquisite penmanship is your forte, then you have missed your century. But even a valued competency may go undeveloped in those who are unwilling or unable to compete. Competitive social structures deemed most fair by those who promote meritocratic selection do not develop valued excellencies in those who prefer more cooperative structures:

> Human nature demands, no doubt, space and elbow-room. But there is an excellence of repose and contentment, as well as of effort; and, happily, the mass of mankind are not all elbows (Tawney, 1952, p. 112).

The meritocratic reformers saw less cooperation and more competition in nature and in society. Early Calvinist dichotomies of elect and damned gave way to the unbroken evolutionary scale of Spencer. This simple reading of evolutionary theory seemed to offer an absolute definition of potential for merit. Those low on the scale of intelligence were no longer indispensable parts of the body politic, but at best "democracy's ballast, not always useless but always a potential liability" (Terman, 1922, p. 658). A more careful reading of Darwin shows greater subtlety:

> The meritocratic selector and the experimental reformer alike missed the point of Darwin's theory. The theory did not posit that <u>generally</u> superior creatures evolve. [Rather, Darwin was]. . . concerned with fitness to survive in a <u>particular</u> ecology. To foster development of a wide variety of persons, then, one must offer a wide variety of environments. A social reform that would standardize the environment

> (whether to fit the average person, or the present elite, or the present proletariat) is inevitably procrustean, conservative, and self-limiting. (Cronbach & Snow, 1977, p. 11)

Messick (1989) argues that test validation must attend to the consequences of test use. These consequences are much broader than the issues of bias or heritability that are typically raised by critics of mental testing. At root is a conception of who we are and how we structure our societies. Mental testing serves the meritocratic ideal. Achievement testers hoped to rank by accomplishment; intelligence testers hoped to rank by potential. Either type of ranking, but especially ranking by potential, can slide into a presumed ranking by worth. One need not read too much in the old literature on intelligence testing or even the recent literature on gifted education to see abundant evidence of this (Margolin, 1994). But for every child who receives the happy news of good heredity or good accomplishment, another receives the unhappy news that -- in a very fundamental sense -- he or she is inadequate.

> In an older society, the humble member . . . can attribute his lowly status to God's will, to the ancient order of things or to a corrupt and tyrannous government. But if a society sorts people out efficiently and fairly according to their gifts, the loser knows that the true reason for his lowly status is that he is not capable of better. That is a bitter pill for any man. (Gardner, 1961, pp. 71-72).

The pill can be even more bitter for a child. In looking back at the wreckage of his own early school days, Churchill observed, "It is not pleasant to feel one self so completely outclassed and left behind at the very beginning of the race" (in West, 1991, p. 149).[9]

Tests emphasize, even magnify human differences. Those who trade in the currency of individual differences thus rarely consider questions of human equality. "Clever men," says Tawney (1952) "are impressed by their difference from their fellows; wise men are conscious of their resemblance to them" (p. 81). Equality takes several forms. Most commonly it refers to political equality, which renders all equal before the law and guarantees all an equal right to participate in and influence government, and to social equality, which negates social class distinctions. It was precisely this sort of political and social equality that most impressed the young De Tocqueville about the America he visited in the 1830's:

> In America, where privileges of birth never existed and where riches confer no peculiar rights on their possessors, men unacquainted with one another are very ready to frequent the same places and find neither peril nor advantage in the free exchange of their thoughts ... their manner is therefore natural, frank, and open. (De Tocqueville, 1945, p. 179).[10]

---

[9] Churchill's case is particularly interesting in that it shows the triumph of ambition, talent, and social class over early school failure. (See Ceci, 1990, for a discussion of the importance of social class over IQ, even for the gifted students in the Terman studies.)

[10] De Tocqueville also believed that religious faith was the ultimate foundation of a democratic society: "This great utility of religions is...more obvious among nations where equality of conditions prevails than among others...Equality, which brings great benefits to the world, nevertheless suggests to men...some very dangerous propensities. It tends to isolate them from one another, to concentrate every man's attention

But such equalitarian attitudes, even if they were not as widespread as De Tocqueville believed, have surely shrunk. Herrnstein and Murray (1994) claim that the culprit is the confluence of ability and wealth in U.S. society that has occurred in the past 30 years. Kaus (1992) claims that economic disparity is but one symptom of the loss of a more fundamental of equality of worth, or what Gardner (1961) calls equality of respect. In its most exalted, usually religious form, this meant that all humans were not simply equal but equally precious, and thus equally worthy of respect no matter what their external circumstances. It was this type of equality -- based as it was on a qualitative difference between man and other animals -- that Darwin's theory seemed to dispel.[11] If humans were not equally fit for survival, then differences in accomplishment and thus in social status need not be as capricious as egalitarian reformers believed.

> What really bothers liberals about American society? ... Is it that a whole class of Americans -- mainly poor, black Americans -- have become more or less totally isolated from the rest of society, and are acquiring the status of a despised foreign presence? Is it that the wealthiest 20 or 30 percent are 'seceding,' as Robert Reich puts it, where they rarely even meet members of non-wealthy classes, except in the latter's role as receptionists and repairmen? (Or)... is it the gnawing sense that in their isolation, these richer Americans not only are passing on their advantages to their children, but are coming to think that those advantages are deserved, that they and their children are, at bottom, not just better off but better? (Kaus, 1992, p. 17)

If so then they are following a path well trod by several generations of theologians, philosophers, biologists, psychologists, and psychometricians. Unlike their counterparts at the turn of this century, however, current mental testers generally try to avoid entanglements in public policy. Those who bravely step out of the ivory tower to challenge public opinion are rarely prepared for the wintry blast that greets them (see Cronbach, 1975). But hiding from the policy debates does not help us perform our task better. We have thought about differences; we need to think again about equality, and about the unintended as well as the intended consequences of our tests.

> There is again uneasiness about school sorting that shapes life chances. But philosophical issues remain unvoiced while loaded, overblown issues such as heritability get all the attention. Society needs to think once again about the kind of equality it would prefer and about the desired relation between productivity, social status, and standard of living. Most of all, it needs to distinguish between education as preparation for service to society, education as preparation to get more out of living, and education as a means of certifying social status. (Cronbach, 1975, p. 11).

We also need to think not only about the way testing contributes to the discovery of talent and the fair allocation of opportunity, but also to the belief that, as Kaus (1992) put it, that some are not just better off, but better.

---

upon himself; and it lays open the soul to an inordinate love of material gratification. The greatest advantage of religion is to inspire diametrically contrary principles." (1945, p. 23)

[11] Richards (1987) claims, however, that this "received view of Darwinian man (as a completely material being), though potent in modern culture, does not resemble the image shaped by Darwin, Spencer, and the Darwinians writing in the last part of the nineteenth century" (p. 504).

Philosophy of Science

The construct of intelligence as innate ability was firmly rooted in the Zeitgeist of the period during which the first tests were developed. But scientists also had beliefs about the scientific enterprise. Indeed, beliefs about how knowledge is acquired and how conflicts among competing explanations are resolved form core assumptions of methods of inquiry of a discipline at a particular point in time. Collectively, these methods define (or assume) a particular philosophy of science. The question of how competing claims are arbitrated is a somewhat narrower issue of epistemological values, and will be discussed as such.

Positivism versus Realism. Logical positivism was the dominant philosophy of science during the late 19th century when the foundations of modern psychology were laid (Koch, 1959). Positivism is often distinguished from an older view of science, realism. Proponents of realism hold that the methods of science allow direct access to reality. Scientific explanations, in this view, describe the world as it really is. Positivism is a more moderate position. According to this view, scientists form models or theories of the world based on observed regularities. Although constructs and the laws that relate them usefully explain these regularities, they are not necessarily real. However, as Slife and Williams (1995) note, it is difficult for positivists not to take the next conceptual step and begin to believe that the constructs formed to explain regularities in the world (e.g., gravity or intelligence) are in some sense real. From Spearman to the present, those who report factor analyses of correlations among tests have routinely slipped from careful statements about factors representing convenient "patterns of covariation" or "functional unities" to entities that exist in some concrete fashion in the brains of those who responded to the tests. Cognitive psychologists can be even less careful about reifying their constructs.[12] Those steeped in the information-processing metaphor who have enjoined the debate about the meaning of intelligence have invoked limitations in the capacity of working memory, or speed of information processing to explain observed differences in performance. But "working memory" is a construct, not a thing; and information-free mental processes are no more than convenient fictions.

Like the earlier realists, positivists put great emphasis on observation. It is hard to read Wolf's (1973) biography of Binet or Joncich's (1968) biography of E. L. Thorndike (appropriately titled The Sane Positivist) without feeling some of the enthusiasm both felt for observation and experimentation. Application of the methods of science to human behavior promised to ameliorate many social ills. There was no worry about the extent to which theory contaminated or shaped their observations, since, in their worldview, facts existed independent of and prior to theory. Nor was there a concern whether the methods of science, rather than providing an avenue for the discovery of truth, might actually presume a certain set of beliefs about the world.

The logical positivism of the turn of the century has been replaced by a less comforting set of philosophies. One of the more important contributions was made by

[12] Some of the new cognitive scientists, shielded from the annoyance of testing their theories on people, seem particularly susceptible to this disease. I once was engaged in a (what else?) electronic discussion with a new Ph.D from one of the premier AI programs. I remember my astonishment when I realized that he really thought production rules to be entities that existed in peoples' heads, much like they existed in his computer programs!

Popper (1963) when he pointed out the logical asymmetry of proof and disproof. A thousand confirming instances does not prove the statement "All swans are white"; but one instance of a black swan disconfirms the statement. The implication is that theories can not be proven correct, only disproven. But Popper's science was still evolutionary; progress came in a thousand small attempted refutations. Kuhn (1970), on the other hand, argued that progress in science was more often discontinuous. A new paradigm would revolutionize thinking in a domain. An even more extreme view is taken by social constructivists (scientific "constructs" are simply shared understandings within a particular community in a particular culture at a particular point in time) and their postmodernist allies (there is no way to secure knowledge of a universal and objective reality; rather, knowledge is contextual and constructed through social and linguistic convention).

Social scientists trained in the first half of this century seemed often to believe their task was to put forth a good theory and then defend it against all attacks. Those trained after Popper and Kuhn are more likely to see their task differently. For example, Anderson (1983), in the preface to a book describing a theory of cognition, announced his intention to "break" the theory. The increasingly widespread acceptance of the notion of intelligence as a cultural construct is grounded in an even more constructivist philosophy.

Epistemological issues. A central epistemological issue for scientists is how to choose among competing explanations or theories: Should it be parsimony? utility? meaningfulness? perceived truth value? Test users have generally opted for utility. In America, psychological meaningfulness prevailed over parsimony in the theoretical debate at least until the 1960's. Then parsimony reasserted its challenge. This epistemological struggle underpins the most enduring debate about intelligence: is it one or many things? The controversy has a long history, and promises to have an even longer one. In large measure this is because the debate is not only about evidence but also about value, such as whether parsimony, utility, or psychological meaningfulness should be given priority. Hierarchical theories offer a compromise, but, as Vernon (1973) pointed out, may better meet statistical than psychological criteria. By emphasizing parsimony over psychological meaningfulness, such theories have enhanced the status of broad factors such as fluid intelligence (Gf), crystallized intelligence (Gc), and spatial visualization (Gv), and diminished the status of narrower factors, such as most of Thurstone's (1938) primary abilities, and all of Guilford's factors. This may or may not be a good thing. Certainly there is less tendency to attribute effects to special ability constructs that could more parsimoniously be attributed to general ability. However parsimony is only one of several criteria that may be used to arbitrate such decisions. Psychological meaningfulness is perhaps equally important, but has been given less weight of late. Indeed, one could argue that psychological clarity declines as factor breadth increases. In other words, the broadest individual difference dimension -- although practically the most useful -- is also psychologically the most obscure. There has never been the sort of controversy over the meaning of factors such as verbal fluency or spatial ability that routinely attends discussion of G (see Lohman & Rocklin, 1995). On the other hand, tests of narrower abilities have never fared as well as tests of broader abilities when utility was the criterion. It is unlikely that new tests will fare better, in spite of the fact that they are more firmly grounded in theory than many of the older classics. Nevertheless, newer tests (such as the Woodcock-Johnson - Revised) are a boon

for researchers, and may someday show utility as aptitude variables that interact with instructional or other treatment variables.

Summary

The theme of this section is that theories of human intelligence reflect not only good research and good thinking, but also the personal proclivities and professional experiences of theorists, their beliefs about science and how it should be conducted, and by even larger social, religious, and political themes that form the fabric of the cultures in which they lived. None of us are somehow above or outside of these influences. It therefore behooves us to discover our hidden biases and beliefs. But there is a larger picture than the self. Those who have reflected thoughtfully on the social sciences can help us see some of the larger tides in the explanations that we offer, such as the infatuation with, rejection of, and rediscovery of biological explanations of human differences (see, e.g., Degler's, 1991, excellent summary).

In a delightful essay, Bannister (1966) points out a central problem for psychologists is not so much how to explain the behavior of others, but how to account for the fact that the theories the psychologist proposes must also account for the psychologist's own behavior, including the behavior of proposing theories about other people's behavior. "If we are going to make so bold as to utter such statements as 'thinking is a matter of A and B and a little C,' then such statements should equally subsume the thinking which led to them. In like manner, those make statements about human intelligence must constantly remind themselves not only of the limits imposed by their own cognitive competence, but of the inevitable distortions introduced by affect. Thus, one must combine Sternberg's (1990) insight that, for theories of intelligence, the model (or root metaphor) is the message, with Damasio's insight that, for better and for worse, emotion and affect are enmeshed in reason's networks. And although none of us can escape from the limits of the human cognitive system, it helps to be aware of these limits. If nothing else, such awareness encourages a healthy skepticism of our own theories and a greater respect for those with whom we disagree.

Theme III: Psychometrics vrs psychology

Tests must meet multiple, often conflicting standards of excellence. Standards that are clear and quantifiable tend to be enforced more rigorously than standards that are vague and not quantifiable. From a handful of simple assumptions about how latent scores and errors combine to produce observed scores (Spearman, 1904b), an elegant , elaborate, and complex psychometric theory has been developed over the years. Those who would master this theory -- and ancillary developments in multivariate statistics (particularly factor analysis) and scaling -- must devote many years to its study. Few have time to develop the level of sophistication in psychological theory that Spearman or Thorndike or Thurstone also achieved. And even if psychometric and statistical theory took no longer to master now than it did at the turn of the century, psychological research and theory have grown exponentially in the interim. There is thus an increasing conflict between the psychometrics of intelligence tests and the psychology of human intelligence. Psychometricians sometimes hold naive beliefs about learning and cognition (Shepard, 1991), protestations to the contrary notwithstanding (Cizek, 1993).

And experimental psychologists are often even more uninformed about psychometrics.[13] Because of this, the topic of this section could be included as a subheading under the general heading of beliefs about the nature of mental tests, and thereby treated as a part of the previous section. However, the issues are sufficiently unique to warrant separate treatment. At the simplest level, the conflict is between those who emphasize the statistical properties of test scores and those who emphasize their psychological meaningfulness. The conflict has actually been more of a rout, since at every turn those who have advocated statistical theories of mental tests have won the battle. Inevitably, though, psychological assumptions about the meaning of test scores must be made and defended. As Cronbach (1990) put it:

> Sooner or later every tester has to go behind the experience table and behind the test content, to say what processes seem to account for the responses observed. (p.159).

By that point, however, the test has been so molded by psychometric principles that it is often difficult -- and sometimes nigh near impossible -- to wrest psychological meaning from the scores. Explaining why this is so requires a brief reconsideration of the history of ability measurement.

How versus how much

In a poem that once was committed to memory by most American school children, Robert Frost tells how, in looking back on his life, "two roads diverged in a wood" early in the journey. That he took one road rather than the other "made all the difference" in the course of his life. Early in this century, two roads -- two approaches -- for developing intelligence tests diverged. Few seemed to notice at the time, but with hindsight it is clear that the path taken was not the only option. One of the few who seemed to notice was E. L. Thorndike:

> All scientific measurements of intelligence that we have at present are measures of some product produced by the person or animal in question, or of the way in which some product is produced. A is rated more intelligent than B because he produces a better product, essay written, answer found, choice made, completion supplied or the like, <u>or produced an equally good product in a better way, more quickly or by inference rather than by rote memory, or by more ingenious use of the material at hand</u>. (Thorndike et al., 1926, p. 11-12, emphasis added)

The crucial difference here is between tasks that allow inferences about ability from <u>how many</u> items are solved and tasks that allow inferences about ability from <u>how</u> items are solved. The first approach is most informative if ability can be described with a trait

---

[13] On several occasions when Snow and I were writing a long monograph on implications of cognitive psychology for educational measurement, Snow remarked how it might be even more fruitful to turn the problem around and write about the implications of educational measurement for cognitive psychology. It is unlikely, though, that very many cognitive psychologists would attend to such a document. In attempting to explain why experimental psychologists have generally ignored the link between psychophysics and mental tests, Guilford (1954) observed that "the experimental psychologist has been very slow in realizing that he uses mental tests as measuring instruments. He has associated mental tests with individual differences, failing to recognize that they also measure 'occasion differences' in the same individual." (p. 4).

model in which individuals vary in their location along a common scale. Number of items solved, or some transformation of this score, typically defines the scale. Growth -- if it occurs at all -- involves quantitative rather than qualitative changes. The second approach is more informative if abilities can be described by stage-like models. Growth involves qualitative changes in how tasks are solved, not simply whether they can be solved. Although one can construct a continuous scale from such data, interpretation is not the same at all points on the scale. The stage approach is now common among developmental psychologists. But it was once a viable approach for the assessment of intelligence. In one of the better-known early summaries of intelligence testing, Freeman (1926) claims that many early attempts to measure intelligence (including the Binet scale of 1905) "did not emphasize the objective score which the child made so much as his general behavior and the way in which he went about the tasks which were set (before) him." (p. 108, italics added). Indeed the notion of mental age as a continuous variable (as Terman envisioned it) was foreign to Binet's views. The term Binet used was "niveau intellectualle", or "intellectual level." Furthermore, children at the same intellectual level might behave in quite different ways. His 1903 book, L'Etude Experimentale de l'Intelligence (The Experimental Study of Intelligence), "primarily dealt with qualitative differences in personality or mental functioning" (Fancher, 1985, p. 65), not incremental improvement along a common scale. Even Wilhelm Stern, who translated Binet's niveau intellectualle into intelligenzalter or "mental age," was quick to argue that individuals who exhibited the same mental age (or "teleological" intelligence) might approach situations quite differently, thereby exhibiting "phenomenologically" distinct intelligences.

The stage-like view of intellectual development is also evident in the type of task Binet found most congenial for the measurement of intelligence. "We place [this task} above all others, and if we were obliged to retain only one, we should not hesitate to select this one" (Binet & Simon, 1908/1916, p. 189). What was the task? The child was shown three pictures, one at a time, and simply asked "What is this?" or "Tell me what you see here." Three types of responses were distinguished: (a) an enumeration response ("a man, a cart, ...") (b) a descriptive response ("There is an old man and a little boy pulling a cart."), and (c) an interpretive response ("There is a poor man moving his household goods.") Note that all children are shown the same stimuli; intellectual level is inferred from the type of response given rather than from the number of pictures correctly identified. This was the sort of task favored by Piaget (who studied in the lab Binet founded) and other developmental psychologists. The theories of intelligence that emerged from reflections on children's responses to such tasks were rich in process and description, but generally ignored psychometric properties of test most emphasized by those who used tests to rank order individuals. Put another way, the concern of the developmentalist was for understanding what intelligence is and how it develops rather than for identifying more and less intelligent individuals. Those concerned with the latter issue understandably preferred tests in which the examiner (or scorer) merely had to judge whether examinees gave (or chose) a keyed response. Judgments about process are much more difficult to make and to defend.

But this was not simply an issue of the reliability of examiner judgments. The two approaches make quite different assumptions about the nature of the measurement scale. Those who believed that mental measurements could be modeled after physical measurements adopted trait models of ability. Because of this, they were quick to

question the meaningfulness of scales that appeared to be measuring something different at the high end of the scale than at the low end. Although several critics of the Binet scale had raised the issue, it was Robert Yerkes who was the most vocal critic. Yerkes and his associates (Yerkes & Anderson, 1915; Yerkes, Bridges, & Hardwick, 1915) proposed an alternative they called the "point" scale. Although Binet sometimes used the same task at more than one age, a variety of different tasks were presented at each level. Yerkes argued that because of this variety there was no guarantee that the same intellectual functions were required at every age. The Point Scale he proposed consisted of twenty subscales, each containing items of a particular type ordered by difficulty. This was the format used in the Army Alpha and Beta (Yerkes chaired the committee which supervised their development) and their successors (i.e., the Wechsler scales and the homogeneous tests used in factor analytic investigations of abilities). The seemingly reasonable assumption that a test should measure the same thing at all levels thus led to a rather dramatic shift in the type of test administered. Furthermore, because one item is much like the next, such tests appear to be psychologically transparent. There is thus a tendency to substitute a simple labeling of type of response required by the test for a theory of cognitive processing. For example, E.L. Thorndike preferred to describe the intelligence measured by his test by the names of the subtests themselves. The test had four subtasks: Completion, Arithmetic, Vocabulary, and Directions; Thorndike referred to the construct measured by the test as "intellect CAVD." Those who followed the factor-analytic route were only slightly less behavioristic. Factors were generally labeled after an inspection of the content of tests that loaded on them. Although examination of the content of tests is a useful first step in understanding process, it is not a very good nth step. In criticizing a similar practice in achievement testing, Cronbach (1971) observed:

> Professional constructors of achievement tests cross a content outline with a set of response-process categories, expressing the latter in terms such as recall, reasoning, application of principles. Such a specification has value in broadening the test. But, since task operations controlled by the tester are not to be confused with processes used by the subject, it is evident that the usual content-by-process grid is not a universal specification. An item qua item cannot be matched with a single behavioral process. Finding the answer calls for dozens of processes, from hearing the directions to complex integration of ideas. The shorthand description in terms of a single process is justified when one is certain that every person can and will carry out all the required processes save one. Even to speak of 'required processes,' however, is misleading, since the task can perhaps be performed successfully by alternative processes... A proper response specification deals with the result a person is asked to produce, not the process(es) by which he succeeds or fails. 'Reads printed words aloud' is a description of an observable response; it says nothing about (process). (pp. 453-454)

Two further observations. The psychological transparency of ability tests composed of similar items generally reveals very little about the construct itself even though total scores on the test may be good measures of the construct. Consider, for example, the difference between the cognitive psychologist's understanding of the construct "working memory" [as elaborated, for example, in Baddeley's (1986) book-length monograph] and the differential psychologist's understanding of the "memory span" factor (or factors). McNemar (1964) saw things clearly when he observed that there seemed to be no way of even beginning to construct a model of the former from scores on the latter (see Lohman

& Ippel, 1993). Second, even seemingly homogeneous tests often elicit different response strategies from different subjects, or for the same subject on different items (Kyllonen, Lohman, & Woltz, 1984). However, such tasks are rarely constructed in ways that make such variation observable. Thus, important differences about processing strategy routinely go undetected when using tasks not designed to reveal them.

Thus, the adoption of tests that estimated ability by number of items of a particular type correctly solved led to the gradual abandonment of tasks that elicited qualitatively different patterns of responses associated with different levels of mental development. Increasingly sophisticated statistical analyses were performed on the scores derived from these homogeneous tests. Psychological theorizing was relegated to debates about the organization of factors in different models rather than about the nature of intellectual functioning itself. Each new generation of differential psychologists was required to spend an ever larger portion of its graduate training to mastering an ever expanding catalog of statistical methods, under the tacit (and sometimes explicit) assumption that new methods would provide insights old methods could not even approximate. This changed the discipline not only by redirecting the efforts of those interested in solving ill-structured psychological problems to well-structured problems in methodology, but also by attracting to it those more interested in statistical methods than in psychology.

The dominance of methodology over psychology has a subtler aspect. Like a child looking for things to hit with his new hammer, there is a tendency to find problems that can be explored with the newly mastered methodology rather than to find or invent methodologies that best address a psychological issue. Even those who know better find it difficult not to rely on ever refined multivariate methods to carry the burdens of careful experiment and clear thinking.[14]

My own attempts to clarify the measurement of individual differences in speed of response illustrate the problem. One may start with a substantive issue (here, "What is the relationship between speed of processing and spatial ability?"), but quickly become enmeshed in solving one methodological problem after another. Unlike an onion, though, there seems to be no end to the successive layers of this problem. More concretely, I started with a Pearson product moment correlation between response latency and spatial ability and ended up fitting a generalized growth curve to response accuracies [P(C)] on spatial test items presented for varying amounts of time (t):

$$P(C)-.5=\left(\lambda_0+\sum\lambda_i b_i\right)\left(1-e^{\left(\beta_0+\sum\beta_i b_i\right)\left(t-\left[\partial_0+\sum\partial_i b_i\right]\right)}\right)$$

[14] Guttman (1971) offered a similar complaint that measurement had been transmuted into a game of applied statistics. For an example of the over-reliance on methodology, see the paper by Michael, Zimmerman, & Guilford (1950). After proposing several interesting hypotheses about differences in cognitive processing or test demands that might produce different spatial ability factors, the tests were administered with standard directions and the scores factor analysed. Guilford certainly was capable of much more; indeed his Psychometric Methods (1954) is still a classic. And Zimmerman had previously shown how experimental manipulations could alter the factor loadings of a test. Yet here, no variables were manipulated. Even introspective reports were not used in any systematic manner. It seems to have been assumed that a proper factor analysis would sort everything out just fine.

But even this level of complexity does not solve the problem. One must collect so much data to fit this sort of model that practice effects must be modeled as well. Even more troublesome is the fact that the experimental methodology used to collect data in the first place seems to alter how some subjects go about the task. The point to this cryptic tale is that there is no end to it. And the more sophisticated the methodology -- especially if one is not the developer of the methodology but a convert to it -- the less likely it is that one will be aware of all of the assumptions made when using it.

Methodology imposed other changes. Because good factor analyses required many test scores from each examinee, group-administered multiple choice tests replaced individually administered performance tests in the research programs of Thurstone, Guilford, and their students. No longer was the examiner privy to the behavior of individual examinees as they struggled to discover solutions to problems. Is it any wonder that Binet, cramped in a broom closet while administering a series of puzzles to children came to a different understanding of intelligence than those who administered batteries of paper-and-pencil tests to auditoriums of military recruits? Two roads thus diverged in the woods. Developmental psychology took one path; psychometrics the other. Every now and then, thoughtful travelers on the psychometric path would wonder what happened to "process," and whether it was ever possible to infer process from scores on such tests (Cronbach, 1957; McNemar, 1962). Of all the strange ironies that have attended the development of mental tests, none is stranger than the attempt by those of us seeking to develop process theories of human abilities to apply sophisticated methods for detecting individual differences in processing to a class of tasks modeled after tests carefully pruned of such differences. Not surprisingly, some of the most interesting and informative studies of individual differences in process used tasks designed to elicit and reveal qualitative differences in knowledge and strategy [see, e.g., Seigler's (1976) studies of Piaget's balance beam problem].

But there is a further complication. Theories of knowing and learning in Europe and North America may be grouped into three perspectives: behaviorist/empiricist, cognitivist/rationalist, and situative/pragmatist-sociohistoric (Greeno, Collins, & Resnick, 1996). Those who follow the behaviorist/empiricist tradition view knowledge as an organized accumulation of associations and components of skills. E. L. Thorndike's theories of learning and intelligence are a good examples. Those who follow the cognitive/rationalist tradition emphasize that knowledge also consists of concepts and theories, and abilities such as reasoning and problem-solving. Both the constructivist theories of developmentalists such as Piaget and the symbolic information processing theories of cognitive scientists are examples of this tradition. The discussion thus far has emphasized how the construct intelligence would be viewed and measured differently in these two traditions. The third perspective on knowing poses an even greater challenge to traditional comceptions and measures of intelligence. The situative/pragmatist-sociohistoric perspective views knowledge as as distributed among people and the artifacts, tools, books that they use, and the communities and practices in which they participate. On this view, then, intelligence is no longer something that can be disembodied from context, much less something that somehow can be located in the head of the knower.

Implications

The claim is not that intelligence has been mismeasured, or that process-sensitive tasks would make better measures of intelligence, or that a situative view of cognition must guide future assessments. On the contrary, existing intelligence tests efficiently estimate of individual differences in the intellectual competencies required by and developed through formal schooling. But they do not tell us much about what intelligence might be, and thus are not very useful for helping us understand how it develops, or how we might best support that development.[15] These measures were birthed in an era when intelligence was generally thought to be something that could be separated from culture and experience. Indeed, if intelligence is understood not only as a good predictor of success in schooling but also as one of the most important outcomes of formal education (see, e.g., Snow & Yalow, 1982; Husen & Tuijnman, 1991), then one of the most promising functions for intelligence tests is as measures of the extent to which students develop transferable mental competencies. Intelligence tests that would support these sort of inferences would need to start from a different set of assumptions than tests grounded in a selection/prediction model. In particular, they would need to be grounded in a developmental theory of knowledge and skill acquisition and transfer (Lohman, 1993), or in a theory that recognizes the need for harmony between person and situation (see, e.g., Snow, 1995). Although useful for selection and classification, recently developed intelligence tests that are better supported by current theories of ability organization or by some theory of mental processing are no more helpful for the goal of developing abilities than were the old tests. Some simply re-label old subtests and item formats with new "process" names. But even those that avoid this trap are still unwitting captives of Yerkes point scale.

Conclusions

What, then, can we learn from the past that would help us in our quest to understand and measure human intelligence? First and foremost, we can learn that we all see through lenses that are formed by belief and affect. Those who are convinced that the "real" intelligence is innate, or conversely, that all differences between us are caused by culture and experience, need to ask why they find such beliefs congenial. "Follow to its source/Every event in action or in thought" advised Yeats (1949). Those who do this well invariably find more than rational argument at the root of their beliefs about intelligence. Second, we need routinely to attend to the larger social and political issues raised by our attempts to define and measure intelligence. To label someone or something as "intelligent" is to make a value judgment. Such judgments have important social and political -- as well as psychological -- consequences. My limited reading in the history of intelligence testing confirms Samelson's (1979) observation that those most closely allied with intelligence testing were often least able to see these larger issues with much clarity. Thus, we need not only to listen more attentively to those who have considered the broader currents in the history and sociology of ideas, but actively to seek their input. Third, we can learn that much of what we are doing has been done before, which hopefully will enable us to see that new measures of intelligence that are not

---

[15] One could make a stronger claim for the measurement of cognitive styles, i.e., that the use of process sensitive tasks and appropriate measurement models would significantly improve the measurement of style constructs. Here the question clearly is How? rather than How much? Measurement models developed in cognitive psychology to estimate consistencies in strategies are much better suited to the task of measuring "how" (see Lohman & Ippel, 1993; Lohman & Rocklin, 1995).

redundant with the old must either follow new theory (as in Sternberg's attempts to assess "Practical" intelligence) or new methods of assessment.

To come full circle, then, "What can we learn from the history of intelligence testing that might inform future assessments?" Much, but only if we are willing to explore some unfamiliar paths. The view from the well-trodden path comforts more and challenges less than it should. The more difficult question, then, is not whether we can learn, but whether we will learn. Experience says fundamental change is unlikely; hope says it must be possible. Science is grounded not only in belief, but in hope. Therefore, I will hope that, as we embark upon the second century of the scientific study and measurement of human intelligence, we will explore our past more fully, our motives and beliefs more honestly, our psychological theories and psychometric methods more critically, and our options for new ways of conceptualizing, measuring, and developing intelligence more creatively. That is my hope, at least.

References

Ackerman, P. L. (1987). Historical review and appraisal of research on learning, retention, and transfer of human motor skills. Psychological Bulletin, 101, 41-74.

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. Journal of Experimental Psychology: General, 117, 288-318.

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Ayres, L. P. (1909). Laggards in our schools: A study of retardation and elimination in city school systems. New York: Charities Publication Committee.

Baddeley, A. D. (1986). Working memory. Oxford, England: Clarendon.

Bannister, D. (1966). Psychology as an exercise in paradox. Bulletin of the British Psychological Society, 19, 21-26.

Binet, A. (1903). L'etude experimentale de l'intelligence (The experimental study of intelligence). Paris: Schleicher.

Binet, A., & Simon, T., (1916). The development of intelligence in the child. In H. Goddard (Ed.) (E. S. Kite, Trans.) The development of intelligence in children (pp. 181-273). Baltimore: Williams & Wilkins. (reprinted from L'Anee Psychologique, 1908, 14, 1-94).

Brigham, C. C. (1930). Intelligence test of immigrant groups. Psychological Review, 37, 158-165.

Bronfrenbrenner, U. (1979). The ecology of human development. Cambridge, MA: Harvard University Press.

Burt, C. (1909). Experimental tests of general intelligence. British Journal of Psychology, 3, 94-177.

Burtt, E.A., (1939). Types of religious philosophy. New York: Harper & Brothers.

Carroll, J. B. (1987). Jensen's mental chronometry: Some comments and questions. In S. Modgil & C. Modgil (Eds.), Arthur Jensen: Consensus and controversy (pp. 297-307). New York: The Falmer Press.

Carroll, J. B. (1989). Factor analysis since Spearman: Where do we stand? What do we know? In R. Kanfer, P. L. Ackerman, & R. Cudeck (Eds.), The Minnesota symposium on learning and individual differences; Abilities, motivation, and methodology (pp. 43-57), Hillsdale, NJ: Lawrence Erlbaum.

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

Cattell, R. B. (1940). Effects of human fertility trends upon the distribution of intelligence and culture. In G. M. Whipple (Ed.), The thirty-ninth yearbook of the

national society for the study of education. Intelligence: Its nature and nurture (pp. 221-234). Bloomington, IL: Public School Publishing Company.

Ceci, S. J. (1990). On intelligence ... more or less. Englewood Cliffs, NJ: Prentice Hall.

Chapman, P. D. (1988). Schools as sorters: Lews M. Terman, applied psychology, and the intelligence testing movement, 1890-1930. New York: New York University.

Chesterton, G. K. (1922). Eugenics and other evils. New York: Cassell and Company, Ltd.

Cizek, G. J. (1993) Rethinking psychometricians' beliefs about learning. Educational Researcher, 22, 4-9.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. American Psychologist, 12, 671-684.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), Educational measurement (2nd ed., pp. 443-507). Washington, DC: American Council on Education.

Cronbach, L. J. (1975). Five decades of public controversy over mental testing. American Psychologist, 30, 1-14.

Cronbach, L.J. (1990). Essentials of psychological testing (5th ed.). New York: Harper & Row.

Cronbach, L. J., & Schaeffer, G. A. (1981). Extensions of personnel selection theory to aspects of minority hiring (Report 81-A2). Stanford, CA: Stanford University, Institute for Educational Finance and Governance.

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

Cubberly, E. P. (1909). Changing conceptions of education. Boston: Houghton Mifflin.

Damasio, A. R. (1994). Descartes' error: Emotion, reason, and the human brain. New York: Putnam.

Darwin, C. (1859). The origin of species by means of natural selection, or, the preservation of favored races in the struggle for life. London: J. Murray.

Davis, A. (1949). Poor people have brains, too. Phi Delta Kappan, 30, 294-295.

De Tocqueville, A. (1945). Democracy in America: Vol. II. New York: Vintage Books.

Degler, C. N. (1991). In search of human nature: The decline and revival of Darwinism in American social thought. New York: Oxford University Press.

Evans, B., & Waites, B. (1981). IQ and mental testing: An unnatural science and its social history. Atlantic Highlands, NJ: Humanities Press.

Fancher, R. E. (1985). The intelligence men: Makers of the IQ controversy. New York, NY: W.W. Norton.

Festinger, L. (1957). A theory of cognitive dissonance. Evanston, IL: Roco, Peterson.

Fleishman, E. A., & Hempel, W. E. (1954). Changes in the factor structure of a complex psychomotor test as a function of practice. Psychometrica, 19, 239-252.

Frederiksen, N. R. (1981). The real test bias: Influences of testing on teaching and learning. American Psychologist, 39, 193-202.

Freeman, F. N. (1926). Mental tests: Their history, principles and application. Boston: Houghton Mifflin.

Galton, F. (1978). Hereditary genius: An inquiry into its laws and consequences. New York: St. Martin's Press (Orig. work published 1869).

Gardner, J. W. (1961). Excellence: Can we be equal and excellent too? New York: Harper and Brothers.

Gordon, M. M. (1964). Assimilation in American life: The role of race, religion, and national origins. New York: Oxford University Press.

Gould, S. J. (1981). The mismeasure of man. New York: W.W. Norton.

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. In D. Berliner & R. Calfee (Eds.), Handbook of educational psychology (pp. 15-46). New York: Macmillan.

Guilford, J. P. (1954). Psychometric methods. New York: McGraw-Hill.

Guttman, L. (1971) Measurement as structural theory. Psychometrika, 36, 329-347.

Hernshaw, L. S. (1979). Cyril Burt, psychologist. Ithaca, NY: Cornell University Press.

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

Hofstadter, R. (1992). Social Darwinism in American thought. Boston, MA: Beacon Press.

Hunt, J. M. (1961). Intelligence and experience. New York, NY: Ronald Press Company.

Husen, T., & Tuijnman, A. C. (1991). The contribution of formal schooling to the increase in intellectual capital. Educational Researcher, 17, 17-25.

Irvine, S. H., & Berry, J. W. (1988). The abilities of mankind: A revaluation. In S. H. Irvine & J. W. Berry (Eds.), Human abilties in cultural context (pp. 3-59). Cambridge University Press.

Jensen, A. (1982). Reaction time and psychometric g. In H.J. Eysenck (Ed.), A model for intelligence (pp. 93-132). New York: Springer.

Joncich, G. M. (1968). The sane positivist: A biography of Edward L. Thorndike. Middletown, CT: Wesleyan University Press.

Kaus, M. (1992). The end of equality. New York: BasicBooks.

Koch, S. (1959). Psychology: A study of science. New York: McGraw-Hill.

Kuhn, T. S. (1970). The structure of scientific revolutions. Chicago: University of Chicago Press.

Kyllonen, P. C., Lohman, D. F., & Woltz, D. J. (1984). Componential modeling of alternative strategies for performing spatial tasks. Journal of Educational Psychology, 76, 1325-1345.

Laboratory of Comparitive Human Cognition (1982). Culture and intelligence. In R. J. Sternberg (Ed.), Handbook of human intelligence (pp. 642-719). New York: Cambridge University Press.

Lippmann, W. (1976). The abuse of the tests. In N. J. Block & G. Dworkin (Eds.), The IQ controversy: Critical readings (pp. 18-20). New York, NY: Pantheon Books.

Locke, J. (1690/1959). Essay concerning human understanding. New York: Dover.

Lohman, D. F. (1993). Teaching and testing to develop fluid abilities. Educational Researcher, 22, 12-23.

Lohman, D. F. & Ippel, M. J. (1993). Cognitive diagnosis: From statistically-based assessment toward theory-based assessment. In N. Frederiksen, R. Mislevy, I. Bejar (Eds.), Test theory for a new generation of tests (pp. 41-71). Hillsdale, NJ: Erlbaum.

Lohman, D. F., & Rocklin, T. (1995). Current and recurring issues in the assessment of intelligence and personality. In D. H. Saklofske & M. Zeidner (Eds.), International handbook of personality and intelligence (pp. 447-474). New York: Plenum.

Margolin, L. (1994). Goodness personified: The emergence of gifted children. New York: Aldine de Gruyter.

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

Messick, S. (1989). Validity. In R. Linn (Ed.). Educational measurement (3rd ed.). New York: Macmillan.

Michael, W. B., Zimmerman, W. S., & Guilford, J. P. (1950). An investigation of two hypotheses regarding the nature of the spatial relations and visualization factors. Educational and Psychological Measurement, 10, 187-243.

Pick, D. (1989). Faces of degeneration: A European disorder, c. 1848- c. 1918. Cambridge, UK: Cambridge University Press.

Popper, K. R. (1963). Conjectures and refutations: The growth of scientific knowledge. New York: Harper and Row.

Richards, R. J. (1987). Darwin and the emergence of evolutionary theories of mind and behavior. Chicago: University of Chicago Press.

Richards, R. J. (1981). Natural selection and other models in the historiography of science. In M. B. Brewer & B. E. Collins (Eds.), Scientific inquiry and the social

sciences: A volume in honor of Donald T. Campbell (pp. 37-76). San Francisco: Jossey-Bass.

Samelson, F. (1979). Putting psychology on the map: Ideology and intelligence testing. In A.R. Buss (Ed.), Psychology in social context (pp. 103-1168). New York, NY: Halsted Press.

Schama, S. (1987). The embarrassment of riches: An interpretation of Dutch culture in The Golden Age. London: William Collins Sons and Co. Ltd.

Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1981). Task differences and validity of aptitude tests in selection: A red herring. Journal of Applied Psychology, 66, 166-185.

Shepard, L. A. (1991). Psychometricians' beliefs about learning. Educational Researcher, 20, 2-9.

Siegler, R. S. (1976). Three aspects of cognitive development. Cognitive Psychology, 8, 481-520.

Slife, B. D., & Williams, R. N. (1995). What's behind the research? Discovering hidden assumptions in the behavioral sciences. Thousand Oaks, CA: Sage Publications.

Snow, R. E. (199). Abilities in academic tasks. In R.J. Sternberg & R.K. Wagner (Eds.), Mind in context: Interactionist perspectives on human intelligence (pp.). New York: Cambridge University Press.

Snow, R. E., & Yalow, E. (1982). Education and intelligence. In R.J. Sternberg (Ed.), Handbook of human intelligence (pp. 493-585). Cambridge, England: Cambridge University Press.

Spearman, C. E. (1904a). "General intelligence" objectively determined and measured. American Journal of Psychology, 15, 201-293.

Spearman, C. E. (1904b). The proof and measurement of association between two things. American Journal of Psychology, 15, 72-101.

Spearman, C. E. (1923). The nature of intelligence and the principles of cognition. London: Macmillan.

Spearman, C. E. (1930). Autobiography. In C. Murchison (Ed.), A history of psychology in autobiography (Vol 1., pp 299-334). Worcester, MA: Clark University Press.

Spencer, H. (1897). The principles of psychology. New York: Appleton and Company.

Stephan, N. (1985). Biology and degeneration: Races and proper places. In J. E. Chamberlin & S. L. Gilman (Eds.), Degeneration: The dark side of progress (pp. 97-120). New York: Columbia University Press.

Sternberg, R. J. (1985). Beyond IQ: A triarchic theory of human intelligence. Cambridge, England: Cambridge University Press.

Sternberg, R. J. (1990) Metaphors of mind: Conceptions of the nature of intelligence. Cambridge, UK: Cambridge University Press.

Tawney, R. H. (1952). Equality (4th edition). London: George Allen and Unwin, Ltd.

Terman, L. M. (1906). Genius and stupidity: A study of some of the intellectual processes of seven 'bright' and seven 'stupid' boys. Pedagogical Seminary, 13, 307-373.

Terman, L. M. (1922). Were we born that way? World's Work, 5, 655-660.

Thorndike, E. L. (1940). Human nature and the social order. New York: Macmillan.

Thorndike, E. L., Bregman, E. O., Cobb, M. V., & Woodyard, E. (1926). The measurement of intellgence. New York: Columbia University, Teachers College.

Thorndike, R. M., & Lohman, D. F. (1990). A century of ability testing. Chicago: Riverside Press.

Thurstone, L. L. (1938). Primary mental abilities. Psychometric Monographs, No. 1.

Tyack, D. B. (1974). The one best system: A history of American urban education. Cambridge, MA: Harvard University Press.

Vernon, P. E. (1973). Multivariate approaches to the study of cognitive styles. In J. R. Royce (Ed.), Multivariate analysis and psychological theory (pp. 125-148). New York: Academic Press.

Watson, J. D. (1980). The Double Helix: A personal account of the discovery of DNA. New York:

Weber, M. (1958). The Protestant ethic and the spirit of capitalism. (T. Parsons, Trans.) New York: Scribner. (original work published 1904)

West, T. G. (1991). In the mind's eye. Buffalo, NY: Prometheus Books.

White, S. H. (1977). Social implications of IQ. In P. L. Houts (Ed.), The myth of measurability (pp. 23-44). New York, NY: Hart Publishing Company.

Wolf, T. (1973). Alfred Binet. Chicago: University of Chicago Press.

Wooldridge, A. (1994). Measuring the mind: Education and psychologyin England, c. 1860- c. 1990. Cambridge, UK: Cambridge University Press.

Yeats, W.B. (1949). "A dialogue of self and soul" in The collected poems of W.B. Yeats. New York: Macmillan.

Yerkes, R. M., & Anderson, H. M. (1915). The importance of social status as indicated by the results of the point-scale method of measuring mental capacity. Journal of Educational Psychology, 6, 137-150.

Yerkes, R. M., Bridges, J. W., & Hardwick, R. S. (1915). A point scale for measuring ability. Baltimore: Warwick & York.