

I.Q.: A Conceptual Deterrent to Revolution in Education

Kenneth Kaye
University of Chicago

A lot of talk about revolution. Schools are going to be made accountable for their effectiveness, children are going to be given whatever attention they need in order to master a set of skills, parents are going to create alternative ways of organizing learning environments. These changes or others equally drastic are certainly needed. But can they occur?

One kind of deterrent to revolution is institutional in nature. Who has the power to change schools? Who stands to gain, who stands to lose? Or more accurately, which groups hope to gain and which groups fear to lose? Politics can stand in the way of progress, or can spur changes which turn out not to be progressive (1).

There is another kind of deterrent to meaningful change that receives almost no attention. This is the conceptual deterrent, the pattern of thought we rarely question, the set of assumptions that limit our vision. Without a revolution in our concepts of learning and human development, changes in educational practices may turn out to be trivial.

One important deterrent to educational progress today is the concept of I.Q. Without debating the merits of such movements as accountability (2), mastery learning

The Elementary School Journal
October, 1973, pp. 9-23

(3), and alternative schooling (4), we can use them as examples of revolutionary thinking that is stymied by our well-entrenched beliefs about I.Q. The basic assumptions behind accountability, mastery learning, and alternative schooling are that children can be taught, that more effective systems of instruction can be found, and that better schooling makes a difference in development.

Assumptions underlying I.Q. are exactly the opposite: that each child acquires knowledge at his own constant rate and that the relative performance of children is not affected by differences in treatment. The two sets of assumptions are clearly incompatible. We cannot logically hold schools accountable for their pupils' achievement if nothing can be done about that achievement. So we must examine the evidence on which the concept of I.Q. is based.

Pioneer work The "intelligence quotient" was invented between 1900 and 1910 to test the hypothesis that each child maintains a steady rate of mental growth all his life. Because "rate of growth" had to be defined in terms of the knowledge acquired by an average child, the original I.Q. was a ratio between the child's mental age (the age of the average child who could answer as many questions as he did on the Stanford-Binet Scale) and his actual age. If this ratio were found to be fairly constant, schools would be able to make sound predictions about the achievement of individual pupils and earlier decisions about how much they would profit from further schooling. What was not expressed,

but certainly contributed to the adoption of the I.Q. tests, was the fact that the tests would justify the status hierarchy of "good" pupils and "bad" pupils. The responsibility for education would be lifted off the school, if indeed it had ever rested there, and placed firmly on the shoulders of the child.

The logical implications of the I.Q. concept were never thought through. If each child really learns at a fixed rate no matter how well the school teaches him, it makes no difference whether he goes to school at all. At an absurd extreme, will it make no difference even if the child is kept locked in a closet? To be able to argue that schooling was necessary, while differences in achievement between children were inevitable, some unknown pioneer invented "potential," the most that can be achieved under favorable circumstances. One still hears the notion that heredity sets limits on potential while experience determines whether a child reaches his potential. In this view, there is such a thing as cognitive deprivation, but no such thing as education beyond the minimum requirement that a child occupy a seat in a school every day. From the point of view of unaccountable schools, this is the perfect rationale; but it is nonsense. There has never been a child who has learned all that he could have learned, achieved anything like his real potential. Exactly as many people perform better than their I.Q.'s predict as perform worse than prediction. By definition, there are as many overachievers as underachievers. Are the overachievers exceeding their potential?

If "potential" had any meaning,

it would have to be defined as a real ability rather than as an ideal limit. If this ability were fixed by age six, there would be no point in having schools, except as day camps. If it were fixed before age six, perhaps determined at conception, nothing could be done for the cognitive development of the preschool child. If this ability were really fixed, however, and schools could do nothing about children's relative success, why bother to measure this "potential"? Why not just let nature run her course?

To confront this paradox was not in the interest of the schools. It was the function of schools to sort children into a hierarchy, and that function was best served by letting the chips fall where they may. When a child's performance was less than his I.Q. predicted, it was his fault not the school's; he was an underachiever. When performance was better than the test predicted, that was the child's fault, too, not the test's. With what scorn we use the word "overachiever"!

Although the correlation between I.Q. and school performance was far from perfect, the stability of I.Q. appeared to support the hypothesis that children had fixed rates of mental growth. "Stability" denotes a high coefficient of correlation between the scores of particular children when tested and retested several years later. Early testmakers (5) regarded this stability as a triumph. There was no recognition of the fact that the tests had been fiddled with to make I.Q. stable over time, even though the underlying assumption—constant learning rates for individual children—was false. The tests had been

standardized by eliminating "unreliable" items, so that only the items that produced stable results remained. The reason I.Q. tests include items on digit span, analogy, and vocabulary is that subtests made up of these items turned out to yield stable scores.

Later testmakers were able to produce an even stabler I.Q., again by fiddling with the test. They found that a child's ranking in his age group fluctuated less than did the ratio between mental and chronological age. The ratio I.Q. was abandoned in favor of something called "deviation I.Q." (6), which is not really a quotient at all. Nonetheless the stability of the fiddled I.Q. score, which for each child tends to vary only within ten points or so in the school years, is often cited as evidence of the constancy of intelligence. It is not.

The paradox about "potential" ability—if there is such a thing as a child's potential, schools have no need to measure it—was avoided in two ways. First, from the beginning of the mental-testing movement there was an enthusiasm for eugenics. Galton and Terman were the strongest advocates of the position that better breeding, not better schooling, was the key to the advancement of man's intelligence. Both men argued that the less intelligent classes bred like rabbits and threatened to overrun their betters (5, 7). At least this position was logically consistent, but it was unsupported by evidence and anyway the educational system was already too well established to be abandoned in favor of eugenics.

A second way of avoiding the paradox about "potential" was the

claim that ability tests could be used to group children into streams or tracks. The best ways of educating smart children, it was asserted, must be different from the best ways of educating stupid ones. There was no implication that the stupid child could improve his class standing, but perhaps every child would learn more in homogeneous classes than in heterogeneous classes. Nature could not be allowed to run her course without guidance.

Did these conclusions follow logically from the I.Q. concept, or from empirical research on tracking? Neither.

Logically, I.Q.'s are inappropriate criteria for the allocation of children to tracks, or the prediction of growth in different environments (8). On the one hand, the tests were constructed so as to include only the abilities that are unaffected by experience (those that led to stable scores). So, although a particular curriculum might make a difference in some kinds of achievement, we should not expect any difference in digit-span memory, vocabulary growth, or in the kinds of school achievement with which I.Q. correlates. This argument forces us again to ask why the prediction and the track placement are needed at all. On the other hand, it is plausible that for any group of children of roughly similar intelligence—high, average, or low—some hypothetical environment exists in which they would learn at a faster rate than in the heterogeneous group. Plausible—but inconsistent with the assumptions of unitary and fixed intelligence underlying I.Q. (9).

Experience either does or does not make a difference.

As we shall see in the following section, ability grouping has survived despite its logical antipathy to the I.Q. myth. The survival of grouping has been a matter of conflict between an established but little-understood construct, I.Q., and an idea too reasonable to be abandoned, the very belief in education. But the revitalization of educational reason will depend on the abandonment of I.Q. as a criterion for placement.

As for empirical research, there has been surprisingly little (10-13). To test the hypothesis that homogeneous grouping leads to better learning would be difficult but not impossible; one would use double-blind controls against Hawthorne and experimenter effects. It has never been done. A related hypothesis, that children randomly placed in high or low streams will approach the expected norms for those streams, has been tested on a small scale with confusing results (14). The hypothesis has never been tested on a large scale. There are a number of indirect studies indicating that the track remolds the child in its own image, but most are inconclusive (10, 11).

Recently it has been shown that streaming does not increase the between-group variance in test scores (12, 13). Although a discussion of these data will require a longer treatment (15), we should say that inequality in terms of variance is not the only issue. Streaming is not necessarily supposed to affect variance; it is supposed to improve the average child's performance (or

even better, every child's performance). There are no data to support this claim. There is nothing to indicate that children in lower streams learn more than they would in mixed classes.

In short, pioneer testmakers and their successors failed to provide any justification for the belief that a second stream or a third stream was ideal for educating children of second- or third-rate potential. Instead, the stream flows toward a first-, a second-, or a third-rate sea, carrying with it nearly any child placed in its current. The only clear result of using I.Q. and other "ability tests" to sort children into different educational environments is that such tests become increasingly good predictors of school achievement: a self-fulfilling prophecy. The concept of "potential" appears intuitively to have some meaning, if it is not examined too rigorously; but its paradoxes were not really resolved by the promise of tracking.

Even if "potential" had presented no difficulties, the concept of "testing" was never confronted honestly. A test is supposed to discriminate among children. If all pupils did equally well or if performance varied randomly from one testing to the next, it would not be a test. Which groups of children a given test discriminates among is controlled, consciously or not, by the person who standardizes the test. The history of I.Q. test construction provides a good example, but the principle also applies to achievement tests and personnel selection tests.

Beginning with the 1937 re-

vision of the Stanford-Binet Scale, I.Q. tests have consciously been standardized so as not to discriminate between boys and girls. On the 1916 Stanford-Binet Scale, girls were superior by a few points at every age. This superiority was done away with by eliminating some of the items on which girls tended to do better. The average scores for girls and for boys are now about 100. The items eliminated were not just those that related to sex roles, but any that happened to be easier for girls—or rather, just enough of these so that the overall test score did not discriminate between the sexes. This was called making the test "sex fair": more fair for boys, less fair for girls. It is obvious that no statement can be made comparing the "real" intelligence of boys and girls.

If the testmakers had wanted to construct a test that was "fair" to rural children in comparison with urban children, they could have done so by using the same procedure, but apparently it did not occur to them. As for social-class differences and racial differences, to eliminate these by a "fair" standardization of the test would have defeated its major purpose. The test was supposed to show that middle-class children were smarter than lower-class children and that whites were smarter than blacks. If it had not done so, no one would have put any credence in I.Q.

Since blacks and whites were found to differ about four times as much as boys and girls on I.Q. tests, the task of constructing a "race fair" or "culture fair" test is more difficult than eliminating dis-

crimination on the basis of sex (16, 17). But it is possible to find a set of items answered equally well by children from the ghetto and children from the suburbs; this was done in a University of Chicago study more than twenty years ago (16). Non-discriminating tests were never put into practice because a test that failed to discriminate between these key groups of children, and to predict future performance, would be of no use to the schools in their appointed task of perpetuating social inequality. So the testmakers settled for using items that they thought ought to be culture fair, even though their tests continued to discriminate. Because the tests were now called "culture fair," psychologists and teachers could assure themselves that they were discriminating fairly.

Just as nothing can be said about the relative intelligence of boys and girls, nothing can be said about class differences or race differences. The presence or absence of such differences need not depend on anything more than the testmaker's decisions about what items to include and what items to exclude. As we shall see, the problem has not been faced any more squarely by modern testmakers than by their predecessors.

It is tempting to dismiss the infancy of education as an age of innocence whose errors resulted simply from a lack of knowledge (18). This illusion vanishes when we discover that many of the leading figures of the period were aware that the notions perpetrated about intelligence were false. Dewey had no use for the concept of mental measurement—and I have tried to

show that no philosopher could believe in education and accept the concept of I.Q.—but unfortunately Dewey's strategy was largely to ignore the testing movement (19, 20). Within the movement Thurstone, whose concern was testing itself and the identification of factors in skilled behavior, maintained a belief in the educability of intelligence and ignored the logical dilemma created by the use of I.Q. tests (21). But there were other leaders who spoke out forcefully against the fallacies of the I.Q., to no avail. Binet rebelled at "the deplorable verdict that the intelligence of an individual is a fixed quantity" (22) and insisted that the school should regard the child's mind as a field to be cultivated more productively: "One increases that which constitutes the intelligence of a school child, namely the capacity to learn, to improve with instruction" (22).

The principal idea of this paper, that the unproved assumptions behind I.Q. tests are fundamentally at odds with the assumption that schools can teach, was expressed incisively by Walter Lippmann fifty years ago in a series of critical essays on the I.Q. concept and its misinterpretation—essays to which Terman replied satirically, dragging out his credentials instead of responding to the argument (23).

As for the issue of the innate superiority of the rich, Galton's errors in reasoning were recognized seventy years ago by scholars who went unheeded (24). The issues arising from attempts to make the tests "sex fair" were discussed clearly by McNemar in his analysis of the 1937 Stanford-Binet revision

(25). More recently, R. L. Thorndike wrote a book reminding his colleagues that "'underachievement' and 'overachievement' really refer to the imperfectness of our predictions" (26). Arthur Jensen eloquently lambasted the Rorschach test; without recognizing the same flaw in I.Q. testing, he described how a clinician uses test results "mainly to bolster his confidence in his own interpretation derived from other sources" (27).

It cannot be said that critics had much influence on the use of I.Q. tests in schools, or on the prevalence of the untested assumptions underlying the I.Q. concept. The critical ideas necessary to destroy our faith in the myth have been present from the beginning. The voices were not heeded, the practices were not abandoned, because the concept of fixed and predetermined intelligence fit perfectly with our society's conception of man and our intended use of schools. We liked to imagine that our society enabled every man of virtue and vigor to advance himself in life. If people typically did not advance from the ghetto (exceptions were heralded as proof that the model was valid), their failure could not have been due to unequal opportunities. It must have been due to innate inevitable differences in mental ability. The American educational system succeeded in promoting the myth while it failed to cultivate the harvest Dewey and Binet had thought possible.

Sophisticated modern work What sort of progress has been made in recent years? We have just come through an era of nonsense about

raising I.Q.'s, nonsense that displayed the most elementary ignorance about what the I.Q. construct is supposed to be. Mental testing began with the hypothesis that rates of learning might be constant for individuals; the ratio I.Q. was a way to test that hypothesis; immediately the tests were fiddled with and the ratio was abandoned in order to make the I.Q. more "reliable." Then it was asserted, and scholars were willing to believe the assertion, that the reliability of this new test must give support to the original hypothesis of constant individual differences in learning ability.

The consequence of this chain of errors is that we have never been able to test the original hypothesis: all claims for constancy are made in terms of I.Q. testing, yet I.Q.'s have a constancy of their own built into them in the process of test construction. This constancy need not have anything to do with the problem of whether individuals' rates of learning are really inflexible and unresponsive to teaching. That problem is the one addressed by Project Head Start and any other program intended to change the educational destinies of groups of children. To use changes in I.Q. as the criterion of effectiveness in such programs, as most Head Start evaluations have done (28), is to guarantee self-defeat. Because test items that led to changes in I.Q. over time were eliminated as the tests were standardized, the items retained on I.Q. tests are the least likely to show an increased score after a program of compensatory education. Using an I.Q. test to evaluate a teaching program of any kind is like trying to

find out whether an unopened carton of milk has turned sour, by weighing it. The test is meaningless for measuring change because constancy is built into it.

This era of attempts to raise I.Q.'s seems to have ended, except for those engaged in the disproof of particular beliefs about retardation (29). It has ended not because its inherent illogic was recognized, but because its proponents were humiliated by their inevitable failure to produce impressive gains in I.Q. That this failure would be taken as evidence of race and class differences in "potential" was foreseeable (30) but inevitable. We are back where we started: with eugenics and ability grouping.

The most recent wave of eugenics arguments has been triggered by Jensen's conclusion that compensatory education failed because of the high heritability of I.Q. (31). Those who have jumped on Jensen's bandwagon have failed to realize that if the assumptions underlying the notion of I.Q. were correct, its heritability should be nearly 100 per cent. The fact that there is some environmental variance in I.Q. gives the lie to the notion of fixed potential. The problem of compensatory education becomes one of finding interventions that work; there is no question that they exist. The relatively high heritability of I.Q. in our present world is a specious argument for eugenics; in a world we could create by finding productive developmental environments for all children, the heritabilities of traits would be different, I.Q. would be meaningless, and eugenics would be unnecessary. As for the racial

differences in mean I.Q., these could be explained entirely by environmental factors even though heritability within each racial group is high (32, 33).

Selective breeding for traits of intelligence, either planned or unplanned, may be inevitable as Richard Herrnstein has suggested (34). But the arguments he uses depend on a belief in the I.Q. construct. Part of the reason a man and a woman are attracted to one another, Herrnstein argues, is the similarity of their levels of intelligence or education. Geneticists call this "assortative mating." The more important I.Q. becomes for success in our modern world, Herrnstein continues, the more people will sort themselves out for marriage on the basis of I.Q. or related traits. Since I.Q. has already been shown to be highly heritable in our society, children are increasingly likely to inherit I.Q. along with their parents' estates.

The premises of Herrnstein's argument are "ifs." If I.Q. continues to be an important determinant of school and occupational success and if heritability remains high—if, in other words, the schools of the future will be as ineffectual as our present schools at outweighing hereditary discrepancies between social classes—Herrnstein's warning is justified. It is equally true, however, that if education were effective, which would require that it see itself as effective and stop prejudicing accomplishments by prior determinations of "potential," one result would be increased social mobility. Another result would be an increase in the intellectual accomplishments of the human species.

No doubt a future historian will regard Herrnstein's analysis as innocent. All I can conclude is that we have failed to make any progress in our thinking since Galton a century ago.

The other recent development in the area of individual differences in school learning has to do with the problem of ability grouping. "Mastery learning" theorists think of children as differing in the amount and the type of instructional attention they require, rather than in their inherent abilities. The instructional requirements vary according to the skills with which the child enters the program. In other words, Johnny and Jimmy do differ measurably, but if schools work more intensively with Johnny he is capable of learning as much as Jimmy. The responsibility for failure is removed from Johnny and his biological ancestors, and placed on the schools. If children can learn to a specified level of mastery, schools can be held accountable.

Although the "mastery group" is reminiscent of the traditional "stream," it differs in at least two significant ways. Streaming has been a matter of segregating children whether or not the school had anything better to do with them in homogeneous than in heterogeneous groups; in mastery learning there must first be developed a curriculum under which children with a particular configuration of skills and styles can be brought to mastery. For the first time, accountability passes beyond the school to the ivy curtain of the research establishment. Secondly, while streaming has typically swept away the whole child and homog-

enized him with the same set of classmates for most or all of the school day (thus undoubtedly facilitating assortative mating), mastery learning is necessarily specific to each subject matter. Johnny might find himself in the same math program as Jimmy though separated in a special reading program. To some extent this has been true all along, but we have so far failed to recognize that it is the only legitimate form of grouping.

The concept of mastery learning is free of most of the logical inconsistencies in earlier thinking, especially the problem of potential. Potentially, Johnny can learn as much as Jimmy; he and his teachers will just have to work a little harder at it. More important, this approach eliminates the need for prior testing and permanent allocation to a stream; in principle, pupils can be grouped by level of achievement or skill mastery instead of by some reified construct of "aptitude." This approach (which has received impetus from the burgeoning number of diagnosable, treatable learning disabilities) does not require streaming except where, in any specific subject, a pupil demonstrates the need for a different or more intensive instructional program (3). Thus it may be possible for the first time to test the twin hypotheses of general intelligence and constant intelligence. The hypothesis of general intelligence predicts that pupils who need intensive treatment in one subject will need it in all subjects; the hypothesis of constant intelligence predicts that the same pupils will fall into this group all their lives.

After referring several times to these as yet untested assumptions of general intelligence (or "g") and constant (or fixed) intelligence, we should add that many scholars and practitioners deny a belief in these myths while affirming the value of I.Q. tests. They argue, for example, that the test is a valid predictor of school success regardless of one's assumptions about the nature of intelligence (35). But this argument is unconvincing because school success may depend partly on the original I.Q. testing, sorting, and labeling. As we have suggested, independent evidence on the stability of performance across the years and across a variety of subtests is unobtainable from I.Q. data because of the way the tests have been constructed. Intercorrelation of subtests and stability of performance were two of the criteria for inclusion of items.

Another defense of current practice also claims open-mindedness on the question of intelligence but respects I.Q. for its diagnostic value. If a child has a low I.Q. his failure in geography is to be expected; if he has a high I.Q. his failure may be due to laziness or xenophobia. One problem with this use of the tests is that the geography test might be a more accurate assessment of the child's ability than the I.Q. test. If I.Q. scores were unaffected by motivation, perhaps they might provide a helpful reference point to compare general cognitive skills with achievement in a particular subject. But all the textbooks and test manuals caution that performance varies with motivation and anxiety. A teacher may just as well believe the results of

the achievement test. Achievement tests include any question the teacher or the curriculum specialist thinks a student should be able to answer. Thus achievement and mastery placement tests do not prejudice the issue with respect to the nature of mental growth, though they may be unfair in other ways.

On closer look, perhaps mastery learning is not so revolutionary after all. Together with accountability, it may be impossible to implement because we expect other things of teachers (discipline and day care) that make it impossible for them to teach (36). In addition there are subtler obstacles built into our conceptual system in education. Mastery learning eliminates the nonsense about potential but fails to confront the paradox inherent in testing. We have still not come to grips with the basic issue of relative academic performance. If the schools devote their energies to seeing that Johnny learns as much as Jimmy, then Jimmy loses his favored position. His parents are not going to stand for that. A backlash such as occurred in suburban communities in response to Head Start will force the schools to intensify their efforts with Jimmy, too. If the least competent children end up mastering skills we formerly thought could be taught only to a few, this is eminently desirable. But it would do nothing to change the relative performance of different children, or even to reduce the discrepancy between the highest and the lowest status groups. Not everyone will agree that education should have these effects, but the lack of discussion of the problem is deplorable.

OCTOBER 1973

The message of the Coleman report is that the distribution of educational resources has little effect on scholastic achievement, which can be pretty well predicted by economic and social advantages alone (12, 37). If the schools perform a function of selecting individuals for roles in a social hierarchy, they do so primarily by justifying a hierarchy that already exists. This function of the school, and the myths about individual differences on which it rests, remain unaffected by the supposedly increasing sophistication of social science.

Revolutions Changes in other aspects of our society are partly to blame for our educational problems. As our world becomes more different from our grandparents' world, for which the present educational system was designed, the schools become less appropriate as a means of socialization (19, 20, 38, 39). The task is to produce adults who will live in a world we can barely imagine. We cannot be sure what skills those adults will need, but we could provide a model for change and adaptation if the school experience itself were developing along with the world outside. Since children relate to the structure of school as if it were the structure of society (40, 41), we ought to provide a dynamic model.

The fact that we think of our society as changing rapidly in every way leads us to believe that education, too, is changing. We gaze into a stagnant pool, but see reflected on its surface the rapid reform in other sectors. Xerox machines, classroom computer consoles, and closed-circuit television do not in them-

selves advance the technology of teaching. Yes, children could be taught, learning could be accelerated, inequities could be remedied; but it is an illusion to think that much progress has been made in this century. It is an illusion, too, that innocence was responsible for the failures of our predecessors. Our predecessors in mental testing believed what they wanted to believe and gathered the evidence that appeared to support their beliefs.

Finding a way to get American education unstuck might be facilitated by a look at non-American education as well as at American non-education. It is surely significant that in the Soviet bloc and China, where schools are seen as the instruments of social and economic progress, the I.Q. chimera is banned as a tool of Western ideology (42). In other sectors of our own society, where capitalism has been allowed to operate relatively freely (in the fields of data processing and medical technology, for example) progress has been rampant. There are necessary constraints on the free operation of capitalism, of course; education is faced with the problem of how to promote creative enterprise while still guaranteeing the education of every child in every community. This is the continual dilemma of capitalism, one that is not solved by taking an institution out of the arena of rational economics, as we have done with education, defense, prisons, and welfare. Perhaps social science needs a new discipline entitled Comparative Planning or Comparative Revolution, to study the relation between the inevitable and the inexorable.

Our argument so far has implied that educational revolution will not come until after educational psychology makes a paradigm shift. Psychology has sold society a dogmatic set of assumptions that preclude beliefs in the educability of children, the potential of curriculum, and the accountability of schools. Unfortunately, theory in social science tends to follow rather than to facilitate social change. Psychologists are not going to resolve the inconsistencies and the contradictions in their theories of intelligence and education until the political climate forces them to do so. Debunking the I.Q. myth is not in the interests of the academy.

Racism, about which we have heard a great deal in the debates surrounding Jensen (31, 43), Eysenck (33), and Shockley (44)—and which can be traced to Galton (7), Terman (5), and other pioneers—is only one aspect of the bias on which our myths about intelligence have been sewn together. Racism is difficult to prove, and the proof will not necessarily lead to better research or scholarship. But motives worse than racism are involved, which have received less attention. Even if there were no issues of race, sex, or social-class discrimination in I.Q. tests, Jensen would still have written his article in 1969 charging that compensatory education had failed, and we would still have had a raging controversy. For Jensen, along with others including Eysenck and Herrnstein (34), responded directly to a growing threat to the credibility of I.Q. testing. Its credibility rested on the assumption that the tests discriminated because groups differed

in innate potential. Criticism that the tests were not "culture fair" really attacked that fundamental assumption. Compensatory education, aimed at altering the academic careers of children who began with low I.Q.'s, threatened the very concept of stable learning ability.

Thus what Herrnstein calls "psychology's most telling accomplishment to date" (34), the whole ability-testing movement, the livelihood of thousands of psychologists, the annual sale of two hundred million standardized tests (45), and the prestige of psychology as a measurement science were all threatened in the 1960's by a growing faith in the efficacy of education. The argument over heritability has at its root the last-ditch defense of an established profession. This charge, no less ugly and no easier to prove than racism, is well worth consideration because it should lead to reexamination of the status of research and debate in education today.

Like academicians, school administrators are slow to change. They are under pressure to follow rather than lead their communities (46). Revolution must take place in the streets before the schools can change their methods and goals. This means that we have to have accountability before we can have educational revolution. Progress will be made inside the school only after it is demanded by forces outside. Unfortunately we can never make the schools accountable so long as they can point to the concept of I.Q. as justification for their failures to teach. If the concept were valid, the problem would be insoluble. Since it is not, we have the

potential for progress.

One thing we can learn from the Soviet revolution, distasteful as it may have been, is that our attitudes about the roles of schools affect our scientific research and theoretical conclusions. In this country the schools have never seriously been asked to redress social imbalance; so the psychologists who fiddled with I.Q. testing came to conclude that differences in intelligence are permanent and heritable. We will have to bring about practical change in the schools and conceptual change in our scientific faculties at the same time. We require a revolutionary period in which major ideas and basic concepts come into confrontation and are tested for their consistency.

My beliefs that science is partly ideological and that it is advanced by confrontation do not, however, imply that American psychology is about to produce its own brand of Lysenkoists. We are finding increasingly that individual differences in behavior, personality, and achievement are related to genetic factors. Furthermore, some genetic factors are bound to produce sex differences and race differences. Our understanding of these factors cannot and should not be slowed by our convictions about the important goals of education. Neither should our educational policies be entrusted to scientists. The advancement of theory in behavior genetics and developmental psychology is largely irrelevant to the decision of Johnny and Jimmy.

The state has, say, twenty-five thousand dollars to spend on the education of these two hypothetical pupils. Bringing them

both to the same level of achievement twelve years hence might require spending twenty thousand dollars on Johnny and five thousand dollars on Jimmy. Dividing the money equally between them might make Jimmy college material and Johnny not. Allocating all the resources to Johnny might enable him to excel a bit, whereas allocating all of them to Jimmy might produce a brilliant scholar. These are political, not scientific decisions. The I.Q. hoax has been a combination of shoddy science and undeserved trust. Better scientific research, however, may lead to similar conclusions about the heritability of individual differences when environment is held constant. Just as the concept of I.Q. should never have been bought by the schools, the more valid conclusions of future research should not be embraced slavishly by those who determine educational policy. If we want to boost the competence of all or of some of our pupils, we can find techniques for doing so. Genetic differences are inevitable, but not inexorable.

One of the recurring images of modern literature is the stockbroker's wife who runs off with a novelist or a ski instructor. That is better than trying another stockbroker. We need major change, not minor change, and not just change for its own sake. We need real change because, to steal Jensen's rhetoric, the American educational system has been tried and it apparently has failed. Our hope, bolstered by reports of the unmanageability of kids in school, is that the consequences of continued failure are now unbearable.

Notes

1. Paul Peterson. "The Politics of Educational Reform in England and the United States." Paper delivered at the American Political Science Association, Washington, D. C., September 5, 1972.
2. Leon M. Lessinger and Ralph W. Tyler (editors). *Accountability in Education*. Worthington, Ohio: Jones, 1972.
3. James H. Block (editor). *Mastery Learning: Theory and Practice*. New York: Holt, Rinehart and Winston, 1971.
4. Robert J. Havighurst and Daniel U. Levine (editors). *Farewell to Schools*. Worthington, Ohio: Jones, 1972.
5. Lewis M. Terman. *The Measurement of Intelligence*. Boston: Houghton Mifflin, 1916.
6. Samuel Pinneau. *Changes in Intelligence Quotient, Infancy to Maturity*. Boston: Houghton Mifflin, 1961.
7. Francis Galton. *Hereditary Genius*. London: Clay and Sons, 1869.
8. Lee J. Cronbach. "Mental Tests and the Creation of Opportunity." Paper delivered at the American Philosophical Society, Chicago, April, 1970.
9. J. McVicker Hunt. *Intelligence and Experience*. New York: Ronald, 1961.
10. Robert Rosenthal and Lenore Jacobson. *Pygmalion in the Classroom*. New York: Holt, Rinehart and Winston, 1968.
11. Janet D. Elashoff and Richard E. Snow. *Pygmalion Reconsidered*. Worthington, Ohio: Jones, 1972.
12. Christopher Jencks. *Inequality*. New York: Basic Books, 1972.
13. Henry Acland. "Ability Grouping in British Primary Schools." Cambridge, Massachusetts: Center for Educational Policy Research, Reprint No. 4, 1972.
14. Richard L. Peterson. "An Experimental Study of the Effects of Ability Grouping in Grades Seven and Eight." Unpublished doctoral dissertation, University of Minnesota, 1966.
15. Kenneth Kaye and Susan R. Kaye. *The I.Q. Hoax*. In preparation, 1973.
16. Kenneth Eells, Allison Davis, and Others. *Intelligence and Cultural Differences*. Chicago: University of Chicago Press, 1951.
17. Martin Deutsch, J. Fishman, and Others. "Guidelines for Testing Minority Group Children," *Journal of Social Issues*, 20 (May, 1964), 138.
18. Benjamin S. Bloom. "Innocence in Education," *School Review*, 80 (May, 1972), 333-52.
19. John Dewey. *Democracy and Education*. New York: Macmillan, 1916.
20. John Dewey. *Education and the Social Order*. New York: League for Industrial Democracy, 1936.
21. L. L. Thurstone. *The Nature of Intelligence*. New York: Harcourt, Brace, 1924.
22. Alfred Binet. *Les Idées Modernes sur les Enfants*. Paris: Flammarion, 1909.
23. Walter Lippmann, Lewis Terman. *New Republic*, 32-33, October-December, 1922.
24. Frank C. Constable. *Poverty and Hereditary Genius*. London: A. C. Fifield, 1905.
25. Quinn McNemar. *The Revision of the Stanford-Binet Scale*. Boston: Houghton Mifflin, 1942.
26. Robert L. Thorndike. *The Concepts of Over- and Underachievement*. New York: Teachers College, 1963.
27. Arthur Jensen. In *Mental Measurements Yearbook*, p. 503. Edited by Oskar Buros. Highland Park, New Jersey: Gryphon, 1965.
28. An exception is Marshall Smith and Joan Bissell. "Report Analysis: The Impact of Headstart," *Harvard Educational Review*, 40 (Winter, 1970), 51-104.
29. Rick Heber and Howard Garber. "An Experiment in the Prevention of Cultural-Familial Mental Retardation." *Proceedings of the Second Congress of the International Association for the Scientific Study of Mental Deficiency*, Warsaw,

- Poland, August-September, 1970. (Manuscript, Rehabilitation Research and Training Center in Mental Retardation, University of Wisconsin, Madison, 1971.)
30. Jacob W. Getzels. "Pre-school Education," *Teachers College Record*, 68 (March, 1966) 219-28.
 31. Arthur Jensen. "How Much Can We Boost I.Q. and Scholastic Achievement?" *Harvard Educational Review*, 39 (Winter, 1969), 1-123.
 32. Richard Light and Paul Smith. "Social Allocation Models of Intelligence," *Harvard Educational Review*, 39 (Summer, 1969), 484-510.
 33. Kenneth Kaye. Untitled review of H. J. Eysenck, *The I.Q. Argument. Adult Education*, 22 (Spring, 1972), 229-33.
 34. Richard Herrnstein. "I.Q." *The Atlantic Monthly*, 228 (September, 1971), 43-64.
 35. The views opposed in this and the following paragraph are widely held, and it would be unfair to mock a particular author. See any standard textbook.
 36. Albert Shanker. "Accountability: Possible Effects on Instructional Programs." In *Accountability in Education*, pp. 66-74. Edited by Leon M. Lessinger and Ralph W. Tyler. Worthington, Ohio: Jones, 1972.
 37. James Coleman and Others. *Equality of Educational Opportunity*. Washington, D. C.: U.S. Government Printing Office, 1966.
 38. James Coleman. "The Children Have Outgrown the Schools," *Psychology Today*, 5 (February, 1972), 72ff.
 39. C. Arnold Anderson. "Education and Social Change," *School Review*, 80 (May, 1972), 433-58.
 40. Joseph Grannis. "The School as a Model of Society," *Harvard Graduate School of Education Bulletin*, 12 (Spring, 1967), 15ff.
 41. Robert Dreeben. *On What Is Learned in School*. Reading, Massachusetts: Addison-Wesley, 1968.
 42. Helen B. Redl (editor). *Soviet Educators on Soviet Education*. New York: Free Press of Glencoe, 1964.
 43. Harvard Educational Review Board (editors). *Science, Heritability, and I.Q.* Cambridge: Harvard University, 1970. (reprint series No. 4)
 44. William Shockley. "Negro I.Q. Deficit: Failure of a 'Social Allocation' Model Warrants New Research Proposals," *Review of Educational Research*, 41 (June, 1971), 227-48.
 45. Marjorie Kirkland. "Effects of Tests on Students and Schools," *Review of Educational Research*, 41 (October, 1971), 303-50.
 46. S. Francis Overlan. "An Equal Chance To Learn," *New Republic*, 166, May 13, 1972, pp. 19-21.