CHAPTER 32

Secular Changes in Intelligence

James R. Flynn

Whether the 20th century has seen intelligence gains is controversial. Whether there have been massive IQ gains over time is not. This difference orders my task. I will (1) describe the range and pattern of IQ gains; (2) discuss their cognitive significance; (3) describe their significance for today's world; (4) argue that they suggest a new theory of intelligence; (5) speculate about what may happen during the 21st century.

The Evidence and Its Peculiarities

Reed Tuddenham (1948) was the first to present convincing evidence of massive gains on mental tests using a nationwide sample. He showed that U.S. soldiers had made about a 14-point gain on Armed Forces tests between World War I and World War II or almost a full standard deviation (SD = 15 throughout). The tests in question had a high loading on the kind of material taught in the classroom and he thought the gains were primarily a measure of improved schooling. Therefore, they seemed to have no theoretical implications, and because the tests were not among those used by clinical psychologists, the practical implications were ignored. It was when Flynn (1984, 1987) showed that massive gains had occurred in the United States on Wechsler and Stanford-Binet IQ tests, and that they had occurred throughout the industrialized world, even on tests thought to be pure measures of intelligence, that IQ gains took center stage. Within a decade, Herrnstein and Murray (1994), the authors of *The Bell Curve*, called the phenomenon "the Flynn effect."

Nations with data about IQ trends stand at 30. Scandinavian nations show that IQ gains may not last much beyond the end of the 20th century, at least in the developed world. Their scores peaked about 1990 and since then, may have gone into mild decline. Several other nations still show robust gains. Americans are still gaining at their historic rate of 0.30 points per year (WAIS 1995–2006; WISC 1989–2002). British children were a bit below that on Raven's from 1980 to 2008, but their current rate of gain is higher than in the earlier period from 1943 to 1980. Other gains cover long periods, so whether the rate varied approaching the present is unknown. Urban Argentines (ages 13 to 24) made a 22-point gain on Raven's between 1964 and 1998. Children in urban Brazil (1930–2002), Estonia (1935–1998), and Spain (1970–1999) made gains akin to the U.S. rate (Colom, Lluis Font, & Andres-Pueyo, 2005; Colom, Flores-Mendoza, & Abad, 2007; Emanuelsson, Reuterberg, & Svensson, 1993; Flynn, 2009a,b,c; Flynn & Rossi-Casé, under review; Must, Must, & Raudik, 2003; Schneider, 2006; Sundet, Barlaug, & Torjussen, 2004; Teasdale & Owen, 1989, 2000).

The developing world shows explosive gains in rural Kenya and the Caribbean. In Sudan, large fluid gains (WAIS Performance Scale) were accompanied by a small loss for crystallized intelligence (Daley et al., 2003; Khaleefa, Sulman, & Lynn, 2009; Meisenberg et al., 2005). If third-world nations continue to gain over the 21st century, and the developed nations do not, the present IQ gap between the two will disappear.

Dutch data illustrate why IQ gains were so disturbing. Between 1952 and 1982, young Dutch males gained 20 IQ points on a test of 40 items selected from Raven's Progressive Matrices (Flynn, 1987). The sample was exhaustive. Raven's was supposed to be the exemplar of a culturally reduced test, one that should have shown no gains over time as culture evolved. These 18-year-olds had reached the age at which performance on Raven's peaks. Therefore, their gains could not be dismissed as early maturation, that is, it was not just a matter that children today matured about two years earlier than the children of yesterday. Current people would have a much higher IQ than the last generation even after both had reached maturity.

These gains created a crisis of confidence: How could such huge gains be intelligence gains? The gains amounted to 1.33 SDs. This would put the average Dutchman of 1982 at the 90th percentile of Dutch in 1952. Psychologists faced a paradox: Either the people of today were far brighter than their parents or, at least in some circumstances, IQ tests were not good measures of intelligence.

Table 32.1 reveals some of the peculiarities of IQ gains. First, it shows how large American gains have been on the most frequently used tests, namely, the Wechsler tests. Both the WISC (Wechsler Intelligence Scale for Children) and the WAIS (Wechsler Adult Intelligence Scale) show full-scale IQ gains proceeding at 0.30 points per year over the last half of the 20th century, a rate often found in other nations, for a total gain of over 15 points. If we link this to earlier data, like that of Tuddenham, the gain over the whole 20th century has been at least 30 points. Second, for children, there is a marked contrast between small gains on subtests close to school-taught subjects (Information, Arithmetic, Vocabulary) and large gains on subtests that require solving a problem on the spot (Picture Completion, Block Design, Coding). The former are often classified as *crystallized* subtests, those that measure what an intelligent person is likely to learn over a lifetime, and the latter as *fluid* subtests, those that measure intelligence by forcing you to solve problems in the test room for which you have no previously learned method.

This WISC pattern of larger gains on fluid than crystallized subtests is international. For example, Raven's gains are huge everywhere and it is the epitome of a fluid test: You study a matrix pattern with a piece missing and must recognize that piece from alternatives, only one of which is correct. For later reference, look at the bottom of the table and note the huge gains on the Similarities subtest, which is a measure of the ability to classify and defies to some degree the crystallized/fluid dichotomy. Also note a new peculiarity that has just come to light. Adults differ from children: The fluid gains of the latter are five times their crystallized gains, while the fluid gains of the former are only slightly greater. This is largely because since 1950, U.S. children have made only a minimal vocabulary gain of 4.40 points, while U.S. adults have made a huge gain of 17.80 points. It is not yet known whether this is an international phenomenon. Other U.S. data suggest that the growing discrepancy between U.S. adults and their children is largely active vocabulary, the words you use, rather than passive vocabulary, the words

Table 32.1. American WISC (Schoolchildren) and WAIS (Adults) Gains

	Rising Full-Scale IQ			
	1947.5	**1972**	**1989**	**2001.75**
WISC	100.00	107.63	113.00	117.63
	1953.5	**1978**	**1995**	**2006**
WAIS	100.00	107.50	111.70	115.07

Contrast between gains on crystallized and fluid subtests (over a shared period of 54 years)

	WISC	*WAIS*
Information (C)	2.15	8.40
Arithmetic (C)	2.30	3.50
Vocabulary (C)	4.40	17.80
Average crystallized	**2.95**	**9.90**
Picture Completion (F)	11.70	11.20
Block Design (F)	15.90	10.25
Coding (F)	18.00	16.15
Average fluid	**15.20**	**12.53**

Subtests ranked by the difference between adult and child gains (over a shared period of 54 years)

	Difference IQ points		*Difference percentages*	
	WAIS – WISC	*Points*	*WAIS/WISC*	*Percentages*
Vocabulary	17.80 − 4.40 =	13.40	17.80 / 4.40 =	405
Information	8.40 − **2.15** =	6.25	8.40 / **2.15** =	**391**
Comprehension	13.80 − 11.00 =	2.80	13.80 /11.00 =	125
Arithmetic	3.50 − 2.30 =	1.20	3.50 / 2.30 =	152
Picture Completion	11.20 − 11.70 =	−0.50	11.20/ 11.70 =	96
Coding	16.15 − 18.00 =	−1.85	16.15/ 18.00 =	92
Similarities	**19.55 − 23.85 =**	**−4.30**	**19.55/23.85 =**	**82**
Block Design	10.25 − 15.90 =	−5.65	10.25/ 15.90 =	64

Sources: Flynn, 2009b; 2009c; under review-b.

you understand when you hear them used (Flynn, under review-b).

The only thing that can be said at present is that the discrepancy does not seem to be because adults have their university education behind them, while their children are still in school. Perhaps it is symptomatic of a trend over the last 50 years for U.S. teenagers to retreat into the subculture of their peers with its own peculiar dialect; and then join the adult speech community as they age and participate in the world of work.

The pattern of IQ gains over time has a final peculiarity, namely, it is not consistently factor-invariant (Wicherts et al., 2004). Factor analysis is a technique that measures the extent to which those who excel on some IQ subtests also excel on others. The tendency toward general excellence is not peculiar to cognitive tests. Just as those who have larger vocabularies also tend to be better at arithmetical reasoning and solving matrices problems, so people who are good at one musical instrument are often good at another, and people good at one sport are often good at almost all sports. The measure of the tendency for a variety of skills to intercorrelate is call *g* (the general intelligence factor). If the top person on one subtest of the WISC topped all the others, and so on down the line, *g* would "explain" 100 percent of the pattern of test performance and have a value of 1.00. If a person's score on each subtest were no more of an indication of their performance on any other subtest than a score chosen at random, *g* would be zero.

One subtest may have a higher *g*-loading than another. This means that it is a better guide as to who will do well on the other subtests. For example, if you added an 11th WISC subtest on shoe tying, it would have a *g*-loading of close to zero: How fast you tie your shoes would have little relation to the size of your vocabulary. On the other hand, your score on the Vocabulary subtest might be a pretty good predictor of your scores on the other subtests (except shoe tying) and get a *g*-loading of 0.75. You could then rank the subtests into a hierarchy according to the size of their *g*-loadings. When this is done, it is evident that the skills with the greatest cognitive complexity top the *g*-loading hierarchy, which is to say that the more complex the task, the greater will be the gap between high-IQ people and the average person. This seems to give *g* a good case to be identified with intelligence and suggests that there might be a latent trait, general intelligence; and that to the extent to which a person possesses that trait, the better he or she will do on a whole range of cognitive tasks.

We can now understand why it is thought significant that IQ gains are not consistently factor invariant. As far as *g* is concerned, this means that when we rank subtests by their *g*-loadings, we find that the magnitude of IQ gains on the various subtest does not tally: The largest IQ gain over time may be on a subtest with an average *g*-loading and the smallest gain may be on a subtest with an above-average *g*-loading. This convinced Jensen (1998) that the bulk of IQ gains were not *g* gains and therefore, were not intelligence gains. He suggests that IQ gains may be largely "hollow," that is, they are a bundle of subtest-specific skills that have little real-world significance.

Two Kinds of Significance

Before we accept the interpretation of IQ gains as hollow, it is useful to supplement factor analysis with functional analysis. Factor analysis may disclose latent traits but no one can do latent traits. What we do in the real world is perform, better or worse, functional activities, such as speaking, solving arithmetic problems, and reasoning about scientific and moral questions. To contrast the two kinds of analysis, I will use a sports analogy.

If we factor analyzed performances on the 10 events of the decathlon, a general factor or *g* would emerge and very likely subordinate factors representing speed (the sprints), spring (jumping events), and strength (throwing events). We would get a *g* because at a given time and place, performance on the 10 events would be intercorrelated, that is, someone who tended to be superior on any one would tend to be above average on all. We would also get various *g*-loadings for the 10 events, that is, superior performers would tend to rise further above average on some of them than on the others. The 100 meters would have a much higher *g* loading than the 1,500 meters, which involves an endurance factor not clearly necessary in the other events.

Decathlon *g* might well have much utility in predicting performance differences between athletes of the same age cohort. However, if we used it to predict progress over time and forecast that trends on the 10 events would move in tandem, we would go astray. That is because decathlon *g* cannot discriminate between pairs of events in terms of the extent to which they are functionally related.

Let us assume that the 100 meters, the hurdles, and the high jump all had large and similar *g* loadings, as they almost certainly would. A sprinter needs upper body strength as well as speed, a hurdler needs speed and spring, a high jumper needs spring and timing. I have no doubt that a good athlete would best the average athlete handily on all three at a given place and time. However, over time, social priorities change. People become obsessed with the 100 meters as the most spectacular spectator event (the world's fastest human). Young people find success in this event a secondary sex characteristic of great allure. Over 30 years, performance escalates by a full SD in the 100 meters, by half a standard deviation in the hurdles, and not at all in the high jump.

In sum, the trends do not mimic the relative *g* loadings of the "subtests." One pair of events highly correlated (sprint and hurdles) shows a modest trend for both to move in the same direction and another pair equally highly correlated (sprint and high jump) shows trends greatly at variance. Factor loadings have proved deceptive about whether various athletic skills are functionally independent. We can react to this in two ways: Either confront the surprising autonomy of various skills and seek a solution by depth analysis of how they function in the real world; or deny that anything real has happened and classify the trends over time as artifacts. The second option is sterile. It is equivalent to saying that if trends are not factor invariant, they are artifacts by definition.

It is better to talk to some athletics coaches. They tell us that over the years, everyone has become focused on the 100 meters and it is hard to get people to take other events as seriously as in the past. They point out that sprint speed may be highly correlated with high jump performance but past a certain point, it is actually counterproductive. If you hurl yourself at the bar at maximum speed, your forward momentum cannot be converted into upward lift and you are likely to time your jump badly. They are not surprised that increased sprint speed has made some contribution to the hurdles because speed between the hurdles is important. But it is only half the story: You have to control your speed so that you take the same number of steps between hurdles and always jump off the same foot. If you told these coaches that you found it surprising that real-world shifts in priorities, and the real-world functional relationships between events, ignored the factor loadings of the events, they would find your mind-set surprising.

Back to the WISC subtests: Arithmetic, Information, Vocabulary, and Similarities all load heavily on *g* and on a shared verbal factor. Despite this, as Table 32.1 shows, between 1947 and 2002, American children gained 24 points on Similarities, 4 points on Vocabulary, and only 2 points on Arithmetic and Information. This is to say that the pattern of gains bears little relation to factor loadings and cannot qualify as factor invariant. However, as usual, factor analysis was done in a static setting with social change held constant. It has no necessary applicability to the dynamic scenario of social priorities altering over time. Thus, *g*-loadings turn out to be bad guides as to which real-world cognitive skills are merely correlated and which are functionally related. To anticipate, a social change over time like people putting on scientific spectacles might greatly enhance the ability to classify (Similarities) without affecting everyday vocabulary or fund of general information. Nonetheless all of these trends would be of great significance, and to dismiss them as "hollow" would be a barrier to understanding the cognitive history of our time.

Interpretation and Causes

Ideally, everyone would approach the cause of massive IQ gains evidentially. But inevitably, a scholar's interpretation of their significance affects his or her list of what causes seem most likely.

If you think that IQ trends are significant as barometers of a shift in cognitive priorities over time, you are likely to focus on cultural factors. But if you believe that they are mainly hollow with a residue that is true intelligence or *g* gains, and that *g* is a latent trait that has its home in brain physiology, you will turn to causes that might affect brain physiology, such as improved nutrition or hybrid vigor (Lynn, 1989, 1990, 1993, 1998; Migronni, 2007). The latter refers to the fact that too much inbreeding is a negative influence on a whole range of human traits including intelligence, as inbreeding between first and second cousins eventually produces IQ deficits. If a nation's population was divided at the beginning of the 20th century into small and inbred communities and then, over time, became more mobile, it would reap the benefits of out-breeding (hybrid vigor) and the nation's mean IQ would rise.

The evidence calls enhanced out-breeding into question as an important cause, at least in developed nations in the 20th century. America was never a collection of isolated communities that discovered geographical mobility only in the 20th century. Right from the start, there was a huge influx of migrants who settled in both urban and rural areas. There were major population shifts during settlement of the West, after the Civil War, and during the World Wars. The growth of mobility has been modest: In 1870, 23% of Americans were living in a state other than the one of their birth; in 1970, the figure was 32% (Mosler & Catley, 1998). Recent data from Norway compare the scores of males as they reach 18 with the scores of their older siblings who reached 18 a few years earlier. If the younger sibling outscores the older, this signals an IQ gain over time (the reverse would signal a loss over time). The IQ trends yielded by these comparisons exactly match the magnitude of the nation's IQ trends (Sundet et al., in press). Because siblings cannot differ in their degree of out-breeding, this shows that hybrid vigor has not been a factor in modern Scandinavia. If it had, the within-sibling estimate would fall short of the actual trend.

In the developed world, better nutrition was probably a factor before 1950, but not since. The nutrition hypothesis posits greater IQ gains in the lower half of the IQ curve than the upper half. The assumption is that even in the past, the upper classes were well fed, while the nutritional deficiencies of the lower classes have gradually diminished. IQ gains have been concentrated in the lower half of the curve in Denmark, Spain, and Norway, but not in Argentina, France, the Netherlands, and the United States. Norway is actually a counterexample: Height gains were larger in the upper half of the distribution while IQ gains were higher in the lower half (Sundet, Barlaug, & Torjussen, 2004). It is unlikely that enhanced nutrition both raises height more than IQ and IQ more than height. British trends are fatal. They do not show the IQ gap between the top and bottom halves reducing over time. The difference was large on the eve of the Great Depression, contracted 1940 to 1942, expanded 1964 to 1971, contacted 1972 to 1977, and has expanded ever since. No coherent dietary history of England can offer the alteration of feast and famine needed to explain these trends (Flynn, 2009a, 2009c).

As noted, those who think IQ trends are barometers that register a shift in cognitive priorities over time will look toward cultural evolution for causes. Flynn (2009a) tried to simplify the explanatory task by focusing on the observation that the largest IQ gains were on Raven's Progressive Matrices and the Similarities subtest of the Wechsler battery.

He asked what "habits of mind" people needed to get the right answers as given in the scoring manuals. Take Similarities: When asked, "What do dogs and rabbits have in common?" the correct answer is that "they are both mammals" rather than "we use dogs to hunt rabbits." The right answer assumes that you are conditioned to look at the world through scientific spectacles – as something to be understood by classification rather than through utilitarian spectacles – as something to be manipulated to advantage. Raven's is all about using logic to deal with sequences of abstract shapes that have no counterpart in concrete reality. If a mind is habituated to taking hypothetical problems seriously and to using logic to deal with the hypothetical, this seems perfectly natural. If you are unaccustomed to using logic for anything but to deal with the concrete world, and indeed distrust reasoning that is not grounded in the concrete, you are unaccustomed to the change of gears that Raven's requires. Like classification, the reasoning rewarded is of the sort that science, which is all about taking explanatory hypotheses seriously, entails.

The next step is rather like an archaeological excavation: Dig into the past hoping to find evidence that appears relevant and assemble it bit by bit. Fortunately, Luria recorded interviews with isolated rural people (Russians in the 1920s) who still lived in prescientific cognitive environments. Here is one about classification:

Fish and Crows (Luria, 1976, p. 82)

Q: What do a fish and a crow have in common?
A: A fish – it lives in water. A crow flies. If the fish just lies on top of the water, the crow could peck at it. A crow can eat a fish but a fish can't eat a crow.
Q: Could you use one word for them both?
A: If you call them "animals," that wouldn't be right. A fish isn't an animal and a crow isn't either. A crow can eat a fish but a fish can't eat a bird. A person can eat a fish but not a crow.

Note that even after an abstract term is suggested, the "correct" answer is still alien. Today we are so familiar with the categories of science that it seems obvious that the most important attribute things have in common is that they are both animate, or mammals, or chemical compounds. However, people attached to the concrete will not find those categories natural at all. First, they will be far more reluctant to classify. Second, when they do classify, they will have a strong preference for concrete similarities (two things look alike, two animals are functionally related, for example, one eats the other) over a similarity in terms of abstract categories. The Similarities subtest assumes exactly the opposite, that is, it damns the concrete in favor of the abstract.

Here is an interview about using logic to analyze the hypothetical:

Camels and Germany (Luria, 1976, p. 112)

Q: There are no camels in Germany; the city of B is in Germany; are there camels there or not?
A: I don't know, I have never seen German villages. If B is a large city, there should be camels there.
Q: But what if there aren't any in all of Germany?
A: If B is a village, there is probably no room for camels.

Today, we are accustomed to detaching logic from the concrete, and say, "of course there would be no camels in this hypothetical German city." The person whose life is grounded in concrete reality rather than in a world of symbols is baffled. Who has ever seen a city of any size without camels? The inhibition is not primarily due to limited experience but rather to a refusal to treat the problem as anything other than concrete. Imagine that the syllogism said there were no dogs in a large German city. The concrete response is that there *must* be dogs in German cities – who would want or be able to exterminate them all? And if one is not practiced in dealing with using logic on hypothetical problems that at least use concrete imagery, what of the hypothetical problems of Raven's that are stated in terms of abstractions with no concrete referent?

Unlike today, when we are bombarded with symbols, the Americans of 1900 had a poverty of experience with such. The only artificial images they saw were drawings or photographs, both of which tended to be representational. Aside from basic Arithmetic, nonverbal symbols were restricted to musical notation (for an elite) and playing cards (except for the religious). They saw the world through utilitarian spectacles:

Their minds were focused on ownership, the useful, the beneficial, and the harmful; and not on the hypothetical and abstract classification.

Genovese (2002) has done his own dig into America's past. He compared the exams the state of Ohio gave to 14-year-old schoolchildren between 1902 and 1913 and between 1997 and 1999. The former tested for in-depth knowledge of culturally valued information; the latter expected only superficial knowledge of such information and tested for understanding complex relationships between concepts. The former were likely to ask you to name the capitals of the (then) 48 states. The latter tended to ask you why the largest city of a state was rarely the state capital (rural members dominated state legislatures, hated the big city, and bestowed the capital on a rural town). Genovese (2002, p. 101) concludes: "These findings suggest that there have been substantial changes in the cognitive skills valued by Ohio educators over the course of the 20th century." We now have a clue as to why there have been virtually no score gains on the WISC general information subtest.

Thus far, the proffered causes of the huge gains on Similarities and Raven's have to do with the minds that took the tests. A full analysis would be multilayered. The ultimate cause of IQ gains is the Industrial Revolution. The intermediate causes are probably its social consequences, such as a better ratio of adults to children, richer interaction between parent and child, better schooling, more cognitively demanding jobs, and cognitively challenging leisure (Neisser, 1998). Donning scientific spectacles with the attendant emphasis on classification and logical analysis is only the proximate cause.

In fairness, biological causes like hybrid vigor and nutrition are usually precise enough to be at risk of falsification. Cultural history, like all history, suggests causes that may be plausible but difficult to quantify and test. More digging is needed if the scenario offered herein is to inspire confidence.

Interpretation and Effects

There is another avenue toward enhanced plausibility. Make "predictions" about what we ought to find in the real world – *if* trends on the WISC subtests are clues to the evolution of functional skills rather than "hollow." Here are a half a dozen: (1) Tutoring children on Raven's should do little to improve their mathematical problem-solving skills. (2) Enhanced performance on school reading and English courses should decline after the age of 14. (3) Enhanced performance in school mathematics should show the same pattern. (4) Popular entertainment should be more cognitively complex and less "literal" in its plot lines. (5) Cognitively demanding games like chess should show large performance gains over time. (6) The quality of moral and political debate should have risen over time.

It is tempting to identify mathematical thinking with the cognitive problems posed by Raven's. Raven's demands that you think out problems on the spot without a previously learned method for doing so, and Mathematics requires mastering new proofs dealing with nonverbal material. They are highly correlated in terms of factor loadings, which seems to signal that they require similar cognitive skills. Therefore, it seems sensible to teach young children Raven's-type problems in the hope that they will become better mathematics problem solvers. U.S. schools have been doing that since 1991 (Blair, Gamson, Thorne, & Baker, 2005, pp. 100–101).

Here IQ gains validate their credentials as a diagnostician of functional relationships between cognitive skills. The large gains on Raven's since 1950 and the virtually nil gains on Arithmetic (see Table 32.1) show that the relationship between the two is no more functional than the relationship between sprinting and the high jump. Sadly, our understanding of the functional process for learning Arithmetic is far behind our understanding of the high jump. Some speculation: Except for mathematicians who link the formulas with proofs, mathematics is less a logical enterprise than a separate reality with its own laws that are at variance with those of the natural world. Therefore, just as infants explore the natural world, children must explore the world of mathematics themselves and become familiar with its "objects" by self-discovery.

Subtests that show minimal gains have as much explanatory potential as those that show huge gains. Since 1950, there have been very minimal gains on the WISC subtests that measure whether children have an adequate fund of general information and a decent vocabulary and whether they can reason arithmetically (Table 32.1). These are very close to school-taught skills. Let us see what they tell us about U.S. trends on the National Association of Educational Progress (NAEP) tests, often called the nation's report card.

The NAEP tests are administered to large representative samples of 4th-, 8th-, and 12th-graders. From 1971 to 2002, 4th- and 8th-graders (average age 11 years old) made a reading gain equivalent to almost four IQ points. However, by the 12th grade, the reading gain drops off to almost nothing (U.S. Department of Education, 2000, pp. 104, 110; 2003, p. 21). The IQ data suggest an interesting possibility. For the sake of comparability, we will focus on WISC trends from 1972 to 2002, rather than on the full period beginning in 1947. Between 1972 and 2002, U.S. schoolchildren made no gain in their store of general information and only minimal vocabulary gains (Flynn, 2009c). Therefore, while today's children may learn to master preadult literature at a younger age, they are no better prepared for reading more demanding adult literature.

You cannot enjoy *War and Peace* if you have to run to the dictionary or encyclopedia every other paragraph. Take Browning's poem:

> Over the Kremlin's pavement bright
> With serpentine and syenite,
> Steps, with other five generals
> That simultaneously take snuff,
> For each to have pretext enough
> And kerchiefwise unfold his sash
> Which, softness self, is yet the stuff
> To hold fast where a steel chain snaps,
> And leave the grand white neck no gash

If you do not know what the Kremlin is, or what "serpentine" means, or that taking snuff involves using a snuff rag, you will hardly realize that these generals caught the czar unaware and strangled him.

In other words, today's schoolchildren opened up an early lead on their parents (who were schoolchildren circa 1972) by learning the mechanics of reading at an earlier age. But by age 17, their parents had caught up. And because current students are no better than their parents in terms of vocabulary and general information, the two generations at 17 are dead equal in their ability to read the adult literature expected of a senior in high school.

From 1973 to 2000, the Nation's Report Card shows 4th- and 8th-graders making mathematics gains equivalent to almost seven IQ points. These put the young children of 2000 at the 68th percentile of their parents' generation. But once again, the gain falls off at the 12th grade, this time to literally nothing (U.S. Department of Education, 2000, pp. 54, 60–61; 2001, p. 24). And once again, the relevant WISC subtest suggests why.

The Arithmetic subtest and the NAEP mathematics tests present a composite picture. An increasing percentage of young children have been mastering the computational skills the Nation's Report Card emphasizes at those ages. However, WISC Arithmetic measures both computational skills and something extra. The questions are put verbally and often in a context that requires more than a times-table-type answer. For example, take an item like this: "If 4 toys cost $6, how much do 7 cost?" Many subjects who can do straight paper calculations cannot diagnose the two operations required: that you must first divide and then multiply. Others cannot do mental arithmetic involving fractions. In other words, WISC Arithmetic also tests for the kind of mind that is likely to be able to reason mathematically.

My hypothesis is that during the period in which children mastered calculating skills at an earlier age, they made no progress in acquiring mathematical reasoning skills. Reasoning skills are essential for higher mathematics. Therefore, by the 12th grade, the failure to develop enhanced mathematical problem-solving strategies begins to bite. American schoolchildren cannot do Algebra and Geometry any better than the previous generation. Once again, although the previous generation was slower to master computational skills, they were no worse off at graduation.

We turn to the worlds of leisure and popular entertainment. Greenfield (1998) argues that videogames, popular electronic games, and computer applications cause enhanced problem solving in visual and symbolic contexts; if that is so, that kind of enhanced problem solving is necessary if we are to fully enjoy our leisure. Johnson (2005) points to the cognitive demands of videogames, for example, the spatial geometry of Tetris, the engineering riddles of Myst, and the mapping of Grand Theft Auto.

However, Johnson's most important contribution is his analysis of television. TV aims at a mass audience and therefore, its level of cognitive complexity is based on an estimate of what the average person can assimilate. Johnson shows convincingly that today's popular TV programs make unprecedented cognitive demands. The popular shows of a generation ago, such as *I Love Lucy* and *Dragnet* and *Starsky and Hutch*, were simplistic, requiring virtually no concentration to follow. Beginning in 1981 with *Hill Street Blues*, single-episode drama began to be replaced with dramas that wove together as many as 10 threads into the plot line. A recent episode of the hit drama *24* connected the lives of 21 characters, each with a distinct story.

Howard (1999) uses traditional games as an informal measure of cognitive gains. He speaks of "cascading feed-back loops": More people want to play chess, the average skill rises, chess clubs form, coaching and chess books improve with rising demand, so you have even better average performance, and so on. He evidences the trend toward enhanced skills by documenting the decline in the age of chess grandmasters. There is no doubt that the standard of play in chess tournaments has risen (Nunn, 1999). Howard makes the same case, although the evidence is less compelling, for feedback loops in other leisure activities that are cognitively demanding such as bridge and go.

Has the quality of political debate risen over the 20th century? Rosenau and Fagan (1997) compare the 1918 debate on women's suffrage with recent debates on women's rights and make an excellent case that the latter shows less contempt for logic and relevance. Note the setting, namely, debate that goes into the *Congressional Record*. That members of Congress have become unwilling to give their colleagues a mindless harangue to read does not mean that all forms of political debate have improved.

We need more research with a proper focus. I suspect that improvement has been limited to written material of some length, that is, material designed to persuade the solitary reader who can take as long as he or she likes to mull over what is said. I anticipate no improvement in two categories. First, speeches to live audiences meant to reduce them to an unthinking mob. William Jennings Bryan's dreadful "Cross of Gold" speech sets the standard for stump oratory today as much as it did over a century ago. Second, there are media events in which the speaker has a few minutes to pack in the most effective sound bites. This is the natural arena of the spin doctor and its standard was set in New Zealand by a candidate who catapulted his party up the polls by using the words "family," "moderate," and "reasonable" more often in five minutes than one would think possible. What we need is a survey covering 50 years of news stories and opinion essays in semiserious publications like *Newsweek* and the *New York Times*.

I know of no study that measures whether the quality of moral debate has risen over the 20th century. However, I will show why it should have. The key is that more people take the hypothetical seriously, and taking the hypothetical seriously is a prerequisite to getting serious moral debate off the ground. When my brother and I would argue with our father about race, and when he endorsed discrimination, we would say, "But what if your skin turned black?" A man born in 1885, and firmly grounded in the concrete, he would reply, "That is the dumbest thing you have ever said – whom do you know whose skin has ever turned black?" I have never encountered contemporary racists who responded in that way. They feel that they must take the hypothetical seriously, and see that they are being challenged to use reason detached from the concrete to show that their racial judgments are logically consistent. The possibility of better moral debate is so important that it too must be subject to systematic investigation.

We can now offer a summary of the real-world implications of IQ gains. Not IQ gains as such, of course, because they have no real-world implications. Rather, it is a summary of the real-world effects of the cognitive trends that IQ scores have registered. Let's take Raven's and the various Wechsler subtests (Table 32.1) one by one:

Raven's: Massive gains show that people have freed logic from analyzing concrete situations to deal with problems put abstractly. This has been a prerequisite for the vast expansion of tertiary education and professional jobs requiring university skills and creative solution of problems on the spot (Schooler, 1998). Taking hypothetical situations seriously may have rendered moral and political debate more reflective. The full potential of this has not been realized because even the best universities do not give their graduates the tools they need to analyze the modern world except perhaps in their area of specialization (Flynn, under review-a).

Similarities: The huge gains mark a transition from regarding the world as something to be manipulated for use to classifying it using the vocabulary of science. This habit of mind is also a prerequisite for higher education.

Performance subtests: Large gains on these are more difficult to interpret. Certainly, the gains on Block Design signal enhanced ability to solve on the spot problems that require more than the mere application of learned rules.

Comprehension: Since 1947, adults have gained the equivalent of almost 14 IQ points and children 11. This subtest measures the ability to comprehend how the concrete world is organized (why streets are numbered in sequence). The greater complexity of life today seems to pose a challenge the average person has risen to meet.

Information: Over 8 points for adults but only 2 points for children. Presumably this reflects the influence on adults of the expansion of tertiary education.

Arithmetic: The small gains here reveal the failure of education on any level to significantly improve arithmetical reasoning.

Vocabulary: A wider gulf exists between parent and child as noted earlier. Serious writers have a larger adult audience able to read their works, although the visual culture of our time may limit the number of those willing to do so.

Another real-world implication of IQ gains: Past standardization samples performed worse than recent ones, and set lower norms. Therefore, obsolete IQ tests give higher scores than up-to-date ones. Therefore, someone who took an obsolete test may get 74 when his or her IQ on current norms would be 69. Since a score of 70 is the cutting line for immunity from the death penalty in America, obsolete tests have literally cost lives (Flynn, 2009b).

Measurement Versus History

The phenomenon of IQ gains has created unnecessary controversy because of conceptual confusion. Imagine an archaeologist from the distant future who excavates our civilization and finds a record of performances over time on measures of marksmanship. The test is always the same, that is, how many bullets you can put in a target 100 meters away in a minute. Records from 1865 (the U.S. Civil War) show the best scoring as 5, records from 1898 (Spanish-American War) show 10, while records from 1918 (World War I) show 50.

A group of "marksmanship-metricians" looks at these data. They find it worthless for measuring marksmanship. They make two points. First, they distinguish between the measure and the trait being measured. The mere fact that performance on the test has risen in terms of "items" correct does not mean that marksmanship ability has increased. All we know is that the test has gotten easier. Many things might account for that. Second, they stress that we have only relative and no absolute scales of measurement. We can rank soldiers against one another at each of the three times. But we have no measure that would bridge the transition from one shooting instrument to another. How could you rank the best shot with a rifle against the best shot with a bow and arrow? At this point, the marksmanship-metrician either gives up or looks for something that would allow him to do his job, perhaps some new data that would afford an absolute measure of marksmanship over time.

However, a group of military historians are also present and it is at this point they get excited. They want to know why the test got easier, irrespective of whether the answer aids or undermines the measurement of marksmanship over time. They ask the archaeologists to look further. If they are lucky, battlefields specific to each time will be discovered. The 1865 battlefields disclose the presence of primitive rifles, the 1898 ones, repeating rifles, and the 1918 ones, machine guns. Now we know why it was easier to get more bullets into the target over time and we can confirm that this was no measure of enhanced marksmanship. But it was of enormous historical and social significance: Battle casualties, the industries needed to arm the troops, and so forth altered dramatically.

Any confusion about the two roles has been dispelled. If the battlefields had been the artifacts first discovered, there would have been no confusion because no one uses battlefields as instruments for measuring marksmanship. It was the fact that the first artifacts were also instruments of measurement that put historians and metricians at cross-purposes. Now they see that different concepts dominate their two spheres: social evolution in weaponry – whose significance is that we have become much better at solving the problem of how to kill people quickly; marksmanship – whose significance is determining which people have the ability to kill more skillfully than other people can. The metrician would not deny that the historian's account is important. The historian has done nothing to undermine what the metrician does. Results on his tests have great external validity. They tell us who is likely to be promoted in each of the three wars (insofar as marksmanship is a criterion) and which of two armies equal in other respects is likely to win a battle (the one with the best marksmen).

I hope this analogy will convince psychometricans (whose job it is to measure cognitive skill differences between people) that my interpretation of the significance of IQ gains over time is not adversarial. Let me make its import explicit.

Some years ago, acting as an archaeologist, I amassed a large body of data showing that IQ tests had gotten much easier over the 20th century in America and elsewhere. Over the century, the average person was getting many more items correct on tests like Raven's and Similarities. The response of intelligence- or *g*-metricians was dual: first, to distinguish IQ tests as measuring instruments from the trait being measured, that is, from intelligence or *g* (if you will); second, to note that in the absence of an absolute scale of measurement, the mere fact that the tests had gotten easier told us nothing about whether the trait was being enhanced. The difficulty was inherent. IQ tests were only relative scales of measurement ranking the members of a group in terms of items they found easy to items they found difficult. A radical shift in the ease/difficulty of items meant all bets were off. At this point, the *g*-metrician decides that he cannot do his job of measurement and begins to look for an absolute measure that would allow him to do so.

However, as a cognitive historian, this was where I began to get excited: Why had the items gotten so much easier over time? Where was the alteration in our mental weaponry that was analogous to the transition from the rifle to the machine gun? This meant returning to the role of archaeologist and finding battlefields of the mind that distinguished 1900 from the year 2000. I found evidence of a profound shift from an exclusively utilitarian attitude to concrete reality toward a much more abstract attitude – to assuming that it was important to classify concrete reality in abstract terms (the more abstract the better); and that taking hypothetical situations seriously had freed logic to deal with not only hypothetical questions but also with symbols that had no concrete referents.

It was the initial artifacts that caused all the trouble. Because they were performances on IQ tests, and IQ tests are instruments of measurement, the roles of the cognitive historian and the *g*-metrician were confused. Finding the causes and developing the implications of a shift in habits of mind over time is simply not equivalent to a task of measurement, even the measurement of intelligence. Now all should see that different concepts dominate two spheres: society's demands – whose evolution from one generation to the next dominates the realm of cognitive history; and *g* – which measures individual differences in cognitive ability. And just as the *g*-metrician should not undervalue the nonmeasurement task of the historian, so the historian does nothing to devalue the measurement of which individuals are most likely to learn fastest and best when in competition with one another.

The direct challenge to those who use conventional IQ tests or the *g* derived from them to measure individual differences comes not from cognitive history but from those who believe they have discovered better measures. No one denies that *g*-loaded IQ tests are useful predictors of things like academic achievement and life outcomes like employment or obedience to the law, and whether children are born in or out of wedlock. However, Sternberg has developed tests that measure creativity and practical intelligence as well the analytic skills emphasized in school, and these may give even better predictions of university marks and job performance (Sternberg, 1988, 2006; Sternberg et al., 2000). Heckman has developed research designs that indicate that noncognitive traits are at least as influential as cognitive traits (Heckman & Rubenstein, 2001; Heckman, Stixrud, & Urzua, 2006).

I have used an analogy to break the steel chain of ideas that circumscribed our ability to see the light IQ gains shed on cognitive history. But an analogy that clarifies one thing can introduce a new confusion. The reciprocal causation between developing new weapons and the physique of marksmen is a shadow of the interaction between developing new habits of mind and the brain.

The new weapons were a technological development of something *outside* ourselves that had minimal impact on biology: Perhaps our trigger fingers got slightly different exercise when we fired a machinegun rather than a musket. But the evolution from preoccupation with the concrete and the literal to the abstract and hypothetical was a profound change *within* our minds that involved new problem-solving activities. Reciprocal causation between mind and brain entails that our brains may well be different from those of our ancestors. It is a matter of use and structure.

If people switch from swimming to weight lifting, the new exercise develops different muscles and the enhanced muscles make them better at the new activity. Everything we know about the brain suggests that it is similar to our muscles. Maguire et al. (2000) found that the brains of the best and most experienced London taxi-drivers were peculiar. They had an enlarged hippocampus, which is the brain area used for navigating three-dimensional space. Here we see one area of the brain being developed without comparable development of other areas in response to a specialized cognitive activity. It may well be that when we do "Raven's-type" problems, certain centers of our brain are active that used to get little exercise; or it may be that we increase the efficiency of synaptic connections throughout the brain. If we could scan the brains of people in 1900, who knows what differences we would see?

So if we can say that the marksman today shoots a *superior gun* to that of his predecessors, can we not say we have a *superior brain* to that of our ancestors? Not superior in every way, of course. The machine gun's gain in firepower is bought at the price of less maneuverability: If someone approaches you from the rear, you would do better to have a rifle that you can turn around in an instant. Our brain may have lost something our ancestors had – something like the wonderful mapping system that Australian Aborigines use in the outback. But, even granting that each generation has a brain adapted to the society of its day, do not our brains deal with an environment of greater cognitive complexity than in 1900? And is that not sufficient reason to say that we are more intelligent?

We can now resolve the question asked at the beginning: Do the huge IQ gains of the 20th century mean we are more intelligent than our ancestors? If the question is, Do we have better brain potential at conception or were our ancestors too stupid to deal with the concrete world of everyday life, the answer is no. If the question is, Do we live in a time that poses a wider range of cognitive problems than those our ancestors encountered, and have we developed new cognitive skills and the kind of brain that can deal with them, the answer is yes. Once we understand what has happened, we can communicate with one another, even if some prefer the label "more intelligent" and others prefer "different." To care passionately about which label we use is to surrender to the tyranny of words.

The Theory of Intelligence

The thesis about psychometics and cognitive history – that they actually complement one another – and the remarks made about the brain imply a new approach to the theory of intelligence. I believe we need a BIDS approach: one that treats the brain (B), individual differences (ID), and social trends (S) as three distinct levels, each having equal integrity. The three are interrelated and each has the right to propose hypotheses about what ought to happen on another level. It is our job to investigate them independently and then integrate what they tell us into a coherent whole.

The core of a BIDS approach is that each of those levels has its own organizing concept and it is a mistake to impose the architectonic concept of one level on another. The best analogy I can find from the history of science is the controversy between Huygens, who championed the wave theory of light, and Newton, who held that light was a stream of corpuscles (particles). Much time was wasted before someone realized that light could act like a wave in certain of its manifestations and like a stream of particles in other manifestations. We have to realize that intelligence can act like a highly correlated set of abilities on one level (individual differences), like a set of functionally independent abilities on another level (cognitive trends over time), and like a mix on a third level (the brain), whose structure and operations underlie what people do on both of the other two levels. Let us look at the levels and their organizing concepts.

Individual differences. Performance differences between individuals on a wide variety of cognitive tasks are correlated primarily in terms of the cognitive complexity of the task (fluid *g*) – or the posited cognitive complexity of the path toward mastery (crystallized *g*). Information may not seem to differentiate individuals for intelligence, but if two people have the same opportunity, the better mind is likely to accumulate a wider range of information. I will call the appropriate organizing concept "General Intelligence" or *g*, without intending to foreclose improved measures that go beyond the limitations of "academic" intelligence.

Society. Various real-world cognitive skills show different trends over time as a result of shifting social priorities. I will call this concept "Social Adaptation." As I have argued, the major confusion thus far has been *either* to insist on using the organizing concept of the individual differences level to assess cognitive evolution, and call IQ gains "hollow" if they are not *g* gains; *or* to insist on using the organizing concept of the societal level to characterize the measurement of individual differences in intelligence, and to deny that some individuals really do have better minds and brains to deal with the dominant cognitive demands of their time.

The brain. Localized neural clusters are developed differently as a result of specialized cognitive exercise. There are also important factors that affect all neural clusters such as blood supply, dopamine as a substance that render synapses receptive to registering experience, and the input of the stress-response system. Let us call its organizing concept "Neural Federalism." The brain is a system in which a certain degree of autonomy is limited by a "higher" organizational structure.

Here I will linger a bit because researchers on this level have the difficult task of explaining what occurs on both of the other two levels. The task of the brain physiologist is reductionist. To illustrate, assume that physiologists have almost perfect knowledge of the brain: When supplied with data on how cognition varies from person to person and from time to time, they can map exactly what brain "locations" underlie the social and life histories supplied. To flesh this out, make the simplifying assumption that the mind performs only four operations when cognizing: classification or CL (of the Similarities sort); liberated logic or LL (of the Raven's sort); practical intelligence or PI (needed to manipulate the concrete world); and vocabulary and information acquisition or VI.

We will posit that the brain is neatly divided into four sectors active respectively when the mind performs the four mental operations, that is, it is divided into matching CL, LL, PI, and VI sectors. Through magnetic resonance imaging scans (MRI) of the brain, we have "pictures" of these sectors. For example, somehow we have MRIs from 1900 that we can compare to MRIs of 2000. When we measure the connections between neurons within the CL and LL sectors, we find that the later brains have thicker connections, and that the extra thickness exactly predicts the century's enhanced performance on Similarities and Raven's.

As for individual differences, we have equally informative pictures of what is going on in the brains of two people in the VI sector as they enjoy the same exposure to new vocabulary. We note that the neurons (and connections between neurons) of one person are better nourished than those of the other due to optimal blood supply (we know just what the optimum is). We note that when the neurons are used to learn new vocabulary, the connections between the neurons of one person are sprayed with the optimum amount of dopamine and the connections of the other are less adequately sprayed. And we can measure the exact amount of extra thickening of the connections the first person enjoys compared to the second. All of this allows us to actually predict their different performances on the WISC Vocabulary subtest.

Given all of the above, brain physiology would have performed its reductionist task: It would have reduced problem-solving differences between individuals and between generations to brain functions; and it would have accommodated both the tendency of various cognitive skills to be correlated on the individual differences level, and their tendency to show functional autonomy on the societal level.

Our Ancestors and Ourselves

IQ trends over time have opened our eyes to a great romance: the cognitive history of the 20th century. Science altered our lives and then liberated our minds from the concrete.

This history has not been written because, as children of our time, we do not perceive the gulf that separates us from our distant ancestors: the difference between their world and the world seen through scientific spectacles. Moreover, because the ability to cope with the concrete demands of everyday life has not been much enhanced, our distant ancestors appear fully human. People use their minds to adapt to the demands of their social environment. Long before the beginning of the 20th century, people felt a strong need to be cognitively self-sufficient in everyday life and long before 1900, virtually everyone who could meet the demands of everyday life had done so. The small percentage that cannot (those who are genuinely mentally retarded) has not varied much over the last 100 years.

Before 1900, most Americans had a few years of school and then worked long hours in factories, shops, or agriculture. Kinship and church provided identity. Slowly society began to demand that the mass of people come to terms with the cognitive demands of secondary education, and contrary to the confident predictions of the privileged, they met that challenge to a large degree. Mass graduation from secondary school had profound real-world effects. The search for identity became a more individual quest. Education created a mass clientele for books, plays, and the arts, and culture was enriched by contributions from those whose talents had hitherto gone undeveloped.

After 1950, the emergence of a new visual culture and perhaps a resistance to the ever increased demands of classroom subjects brought progress to an end in areas like school mathematics and the appreciation of serious literature. Nonetheless, post-1950 IQ cognitive gains have been significant. More and more people continued to put on scientific spectacles. As use of logic and the hypothetical moved beyond the concrete, people developed new habits of mind. The scientific ethos provided the prerequisites for this advance. However, once minds were prepared to attack these new problems, certain social triggers enhanced performance greatly. Post-1950 affluence meant that people sought cognitive stimulation from leisure. It meant that parents had to rear fewer children and they became preoccupied with affording their children a cognitively stimulating environment. Schools became filled with children and teachers less friendly to rote learning, and the world of work offered more and more professional and managerial jobs. These jobs both required and stimulated the new habits of mind. As this last implies, there was causal interaction: New problems developed new skills and better skills allowed us to cope with an even wider range of problems.

The expanded population of secondary school graduates was a prerequisite for the educational advance of the post-1950 era, that is, the huge increase in the number of university graduates. These graduates have gone the farthest toward viewing the world through scientific spectacles. They are more likely to be innovative and independent and therefore, can meet professional and managerial demands. A greater pool of those suited by temperament to be mathematicians or theoretical scientists or even philosophers, more contact with people who enjoy playing with ideas for its own sake, the improvement of managerial efficiency, the enhancement of leisure, the enhancement of moral and political debate – these things are not to be despised.

Quo Vadis

Lynn and Vanhanen (2002) have engendered pessimism by showing that the mean IQs of many nations in the developing world are well below those in the developed world. However, there are signs that IQ gains may cease in developed nations in the 21st century and evidence that they are just taking off in the developing world. These trends would close the developed/developing IQ gap and falsify the hypothesis that some nations lack the intelligence to industrialize. In 1917, Americans had a mean IQ of 70 (against today's norms), which matches the lowest IQs found in the developing world. IQ does not leap from 70 to 100 as a prerequisite for industrial development. The first step toward modernity raises IQ a bit, which paves the way for the next step, which raises IQ a bit more, and so on. The converging IQ trends may be fragile: An environmental crisis might merely inconvenience rich nations while sending poor nations into a downward spiral toward starvation and anarchy.

Despite static IQ, the developed world may enjoy a century of cognitive progress just as exciting as the last 100 years. Science has not only freed logic from the concrete but has also bestowed a second gift, one on which we have not yet capitalized. I refer to a set of wonderful concepts that allow us to critically analyze the modern world: market analysis, basic social science methodology, analytic concepts that make sense of international relations, philosophical progress toward identifying bad argument particularly in ethics, and so forth. But there is no reason for optimism. Universities seem determined to give each graduate one or two of these tools at best. In the larger society, uncritical minds use logic and the vocabulary of science to argue for nonsense (creation science) and fill the schools with confusion. Even universities have become a home to academics that kill critical acumen: those who deny science and reason any special role in the search for truth.

IQ gains over time signal the evolution of minds that can be better educated. They provide no guarantee that the educating will be done.

References

Blair, C., Gamson, D., Thorne, S., & Baker, D. (2005). Rising mean IQ: Cognitive demand of mathematics education for young children, population exposure to formal schooling, and the neurology of the prefrontal cortex. *Intelligence*, **33**, 93–106.

Colom, R., Flores-Mendoza, C. E., & Abad, F. J. (2007). Generational changes on the Draw-a-Man Test: A comparison of Brazilian urban and rural children tested in 1930, 2002, and 2004. *Journal of Biosocial Science*, **39**, 79–89.

Colom, R., Lluis Font, J. M., & Andres-Pueyo, A. (2005). The generational intelligence gains are caused by decreasing variance in the lower half of the distribution: Supporting evidence for the nutrition hypothesis. *Intelligence*, 2005, **33**, 83–92.

Daley, T. C., Whaley, S. E., Sigman, M. D., Espinosa, M. P., & Neumann, C. (2003). IQ on the rise: The Flynn effect in rural Kenyan children. *Psychological Science*, **14**, 215–219.

Emanuelsson, I., Reuterberg, S.-E., & Svensson, A. (1993). Changing differences in intelligence? Comparisons between groups of thirteen-year-olds tested from 1960 to 1990. *Scandinavian Journal of Educational Research*, **37**, 259–277.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, **95**, 29–51.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, **101**, 171–191.

Flynn, J. R. (2009a). Requiem for nutrition as the cause of IQ gains: Raven's gains in Britain 1938 to 2008. *Economics and Human Biology*, **7**, 18–27.

Flynn, J. R. (2009b). The WAIS-III and WAIS-IV: Daubert motions favor the certainly false over the approximately true. *Applied Neuropsychology*, **16**, 1–7.

Flynn, J. R. (2009c). *What is intelligence? Beyond the Flynn Effect*. Expanded paperback edition. New York, NY: Cambridge University Press.

Flynn, J. R. (under review-a). *How to improve your mind*.

Flynn, J. R. (under review-b). *The vocabularies of adults and school children drift apart: WAIS and WISC subtest gains*.

Flynn, J. R., & Rossi-Casé (under review). *IQ gains in Argentina between 1964 and 1998*.

Genovese, J. E. (2002). Cognitive skills valued by educators: Historic content analysis of testing in Ohio. *Journal of Educational Research*, **96**, 101–114.

Greenfield, P. (1998). The cultural evolution of IQ. In U. Neisser, (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 67–79). Washington, DC: American Psychological Association.

Heckman, J. J., & Rubenstein, Y. (2001). The importance of non-cognitive skills: Lessons from the GED testing program. *American Economic Review*, **91**, 145–149.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and non-cognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, **24**, 411–482.

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

Howard, R. W. (1999). Preliminary real-world evidence that average intelligence really is rising. *Intelligence*, **27**, 235–250.

Khaleefa, O., Sulman, A., & Lynn, R. (2009). An increase of intelligence in Sudan, 1987–2007. *Journal of Biosocial Science*, **41**, 279–83.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Johnson, S. (2005). *Everything bad is good for you: How today's popular culture is actually making us smarter*. New York, NY: Rimerhead Books.

Luria, A. R. (1976). *Cognitive development: Its cultural and social foundations*. Cambridge MA: Harvard University Press.

Lynn, R. (1989). Positive correlation between height, head size and IQ: A nutrition theory of the secular increases in intelligence. *British Journal of Educational Psychology*, **59**, 372–377.

Lynn, R. (1990). The role of nutrition in secular increases in intelligence. *Personality and Individual Differences*, **11**, 273–275.

Lynn, R. (1993). Nutrition and intelligence. In P. A. Vernon (Ed.), *Biological approaches to the study of intelligence* (pp. 105–121). Norwood, NJ: Ablex.

Lynn, R. (1998). In support of nutrition theory. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 67–79). Washington, DC: American Psychological Association.

Lynn, R., & Vanhanen, T. (2002). *IQ and the wealth of nations*. Westport, CT: Praeger.

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences*, **97**, 4398–4403.

Meisenberg, G., Lawless, E., Lambert, E., & Newton, A. (2005). The Flynn effect in the Caribbean: Generational change in test performance in Dominica. *Mankind Quarterly*, **46**, 29–70.

Mingroni, M. A. (2007). Resolving the IQ paradox: Heterosis as a cause of the Flynn effect and other trends. *Psychological Review*, **114**, 806–829.

Mosler, D., & Catley, B. (1998). *America and Americans in Australia*. Westport, CT: Praeger.

Must, O., Must, A., & Raudik, V. (2003). The secular rise in IQs: In Estonia, the Flynn effect is not a Jensen effect. *Intelligence*, **31**, 461–471.

Neisser, U. (Ed.). (1998). *The rising curve: Long term gains in IQ and related measures*. Washington, DC: American Psychological Association.

Nunn, J. (1999). *John Nunn's chess puzzle book*. London: Gambit.

Rosenau, J. N., & Fagan, W. M. (1997). A new dynamism in world politics: Increasingly skilled individuals? *International Studies Quarterly*, **41**, 655–686.

Schneider, D. (2006). Smart as we can get? *American Scientist*, **94**, 311–312.

Schooler, C. (1998). Environmental complexity and the Flynn effect. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 67–79). Washington, DC: American Psychological Association.

Sternberg, R. J. (1988). *The triarchic mind: A new theory of human intelligence*. New York, NY: Penguin.

Sternberg, R. J. (2006). The Rainbow Project: Enhancing the SAT through assessments of analytic, practical, and creative skills. *Intelligence*, **34**, 321–350.

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J. A., Wagner, R. K., Williams, W. M., Snook, S. A., & Grigorenko, E. L. (2000). *Practical intelligence in everyday life*. New York, NY: Cambridge University Press.

Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, **32**, 349–362.

Sundet, J. M., et al. (in press). The Flynn effect in sibships: Investigating the role of age differences between siblings. *Intelligence*.

Teasdale, T. W., & Owen, D. R. (1989). Continued secular increases in intelligence and a stable prevalence of high intelligence levels. *Intelligence*, **13**, 255–262.

Teasdale, T. W., & Owen, D. R.. (2000). Forty-year secular trends in cognitive abilities. *Intelligence*, **28**, 115–120.

Tuddenham, R. D. (1948). Soldier intelligence in World Wars I and II. *American Psychologist*, **3**, 54–56.

U.S. Department of Education, Institute of Education Sciences, National Center for Educational Statistics. (2003). *The nation's report card: Reading* 2002, NCES 2003–521, by W. S. Grigg, M. C. Daane, Y. Jin, & J. R. Campbell. Washington, DC: Author.

U.S. Department of Education, Office of Educational Research and Improvement, National Center for Educational Statistics. (2000). *NAEP* 1996 *Trends in Academic Progress*, NCES 97–985r, by J. R. Campbell, K. E. Voelkl, & P. L. Donahue. Washington, DC: Author.

U.S. Department of Education, Office of Educational Research and Improvement, National Center for Educational Statistics. (2001). *The Nation's Report Card: Mathematics* 2000, NCES 2001–517, by J. S. Braswell, A. D. Lutkus, W. S. Grigg, S. L. Santapau, B. Tay-Lim, & M. Johnson. Washington, DC: Author.

Wicherts, J. M., Dolan, C. V., Hessen, D. J., Oosterveld, P., van Baal, G. C. M., Boomsma, D. I., & Span, M. M. (2004). Are intelligence tests measurement invariant over time? Investigating the Flynn effect. *Intelligence*, **32**, 509–538.