



## Overview of the Flynn effect



Robert L. Williams

Clifton, VA 20124, United States

### ARTICLE INFO

#### Article history:

Received 17 August 2012

Accepted 12 April 2013

Available online 6 June 2013

#### Keywords:

Flynn effect

Secular rise in IQ scores

Jensen effect

### ABSTRACT

Following WW2, various researchers found and reported secular gains in IQ, but it was not until additional reports appeared in the 1980s that researchers began to look for the cause or causes. It was quickly apparent that the gains were not limited to any group or nation, but the manifestation of the gains was different depending on time and place. For every discovery, there was a different or opposite result in a different data set. Gains have been large, small, variable, and even negative. Some researchers have found that the gains were on *g*, while more have found no *g* loading. Abstract test formats, such as the Raven have often shown the greatest gains, but gains have also appeared in tests of crystallized intelligence. Some data has shown greater gains for the lower half of the intelligence distribution, while others have shown greater gains in the top half, and others have shown equal gains at all levels. Hypotheses for the causes have included environmental factors, genetic effects, reduced fertility, and methodological dependence. Two models are discussed.

© 2013 Elsevier Inc. All rights reserved.

## 1. Introduction

The secular rise in IQ scores appeared unexpectedly and has defied explanation. Smith (1942) recorded a gain (in Honolulu) over a 14 year span. Later, Tuddenham (1948) found an increased intelligence when he compared inductee scores for the U.S. Army from World War I and World War II and proposed that the gains might be due to increased familiarity with tests; public health and nutrition; and education [the gains from 1932 to 1943 were 4.4 points per decade.]. He cited a high correlation (about .75) between years of education and the Army Alpha and Wells Alpha tests that he was studying.

The secular gain remained relatively dormant until it was rediscovered by Lynn (1982) while working on a comparison of Japanese and U.S. data. It was then rediscovered again, using American data, by Flynn (1984a,b). The raw score gains did not have a name until Herrnstein & Murray (1994) coined the term Flynn effect in their book *The Bell Curve* (p. 307). Some researchers choose to refer to the secular gain as the Lynn–Flynn effect, or use an uppercase FL (Flynn effect) for

the obvious reason that they feel Lynn has been somewhat slighted by not including his name.

Since the early 80s, researchers have found the FE in virtually every group they have examined (Flynn, 1987 and others). They have published a huge number of papers (well over 100) on the gains and possible causes, but the results have been contradictory.

## 2. Gains

FE gains vary from country to country and over different time intervals, but the gains are usually a fraction of a point per year. As a matter of convenience, the gains are usually given as the number of points gained over a decade and written “ $\Delta IQ$ .” A few typical national gains:

U.S.  $\Delta IQ = 3$  (14 points over 46 years, 1932–1978)

Estonia  $\Delta IQ = 1.65$  (12 points over 72 years, 1933/36 to 2006)

Japan  $\Delta IQ = 7.7$  (19 points over 25 years, 1940 to 1965)

Argentina  $\Delta IQ = 6.91$  (21.35 points over 34 years, 1964 to 1998).

E-mail address: [bvv@cox.net](mailto:bvv@cox.net).

[Numerous other rates are given in Flynn and Rossi-Casé (2012).].

South Koreans born between 1970 and 1990 gained at about the same rate as did the Japanese (te Nijenhuis, Cho, Murphy, & Lee, 2012). Chinese gained 4.53 points over 22 years ( $\Delta IQ = 2.1$ ) on the Chinese WPPSI (Liu, Yang, Li, Chen, & Lynn, 2012). FE gains have been found in both industrialized and third world nations. The number of countries showing a FE is subject to change, since additions are frequently reported. Kanaya, Ceci, and Scullin (2005) reported 20 nations; Flynn and Rossi-Casé (2012) reported 31.

Teasdale and Owen (1989) examined two samples of Danish draftees, consisting of 32,862 and 6757 males. They found that the gains were concentrated mostly among the lower IQ levels and concluded that changes in the educational system were driving the score gains. They also performed an interesting test, using Monte Carlo simulations to demonstrate that the FE gain was not caused by a ceiling effect. Flynn and Rossi-Casé (2012) noted that some data sets (they were examining Raven scores) have attenuated SDs because of ceiling effects.

Other researchers, including Lynn and Hampson (1986) and Colom, Lluís-Font, and Andres-Pueyo (2005), have found FE gains that were mainly concentrated in the lower IQ levels. This pattern suggests that the gains are related to improving environmental conditions in non-industrialized countries, rural areas, and low income sectors.

Although it has now been 14 years since Jensen (1998) published *The g Factor*, his discussion of the FE remains current with respect to the items he considered. He reported these U.S. gains:

Raven	$\Delta IQ = 5.69$	
Wechsler	$\Delta IQ = 5.2$	
	Performance	$\Delta IQ = 7.8$
	Verbal	$\Delta IQ = 4.2$

These show greater gains on the most abstract tests and subtests, although it is surprising to see the Wechsler as close to the Raven as the above numbers indicate — both being above the usually cited U.S. rate ( $\Delta IQ = 3$ ).

When Jensen examined subtests more closely, he found that nonscholastic test items showed increases at the same time (same test data sets) that scholastic items were decreasing. He noted that this is not what one would expect to see, but this is indeed what other researchers have reported. Jensen examined the SAT for the period 1952–1990 and found the well known decline. The usual explanation for the decline is that each year more students took the test and most of the additions to the pool of test takers were added below (lower intelligence) the prior group, leading to a decline at the mean. But Jensen corrected for the changes in demographics and showed that 3/4 of the decline was due to the addition of more lower IQ testees, while the remaining 1/4 was a real decline in scores. The  $\Delta IQ$  loss for the SAT was  $-5$  for the time period in question, while the FE gain was  $+3$ . This strongly suggests that the IQ test scores were not reflecting real world gains in intelligence.

### 2.1. Estonia

Thanks to the work done by Olev and Aasa Must, there is a good bit of information about the FE as it has appeared in

Estonia. The messages from their studies are that the FE gains follow different trajectories in different countries and the factors most likely to be driving those changes are also different.

In the Estonian studies, subtests that needed computation skills and mathematical thinking were unchanged over 60 years. The information subtest declined; verbal subtests showed moderate gains; but there were impressive gains in symbol–number and comparison subtests (Must, Must, & Raudik, 2003).

Must, te Nijenhuis, Must, and van Vianen (2009) examined data over a 72 year span and found a relatively small  $\Delta IQ$  of 1.65. But when the eight years from 1998 to 2006 were examined separately, the  $\Delta IQ$  almost doubled to 3 points. The *g* factor loadings were different at the subtest level for each of the three birth cohort groups examined, with the greatest difference between the oldest cohorts compared to the other two relatively recent cohorts.

In recent years, large gains were observed in arithmetic, information, and vocabulary. These gains are opposite from score changes seen in the U.S. and Britain. The authors identified several possible causes: greatly improved education, better nutrition, better health care, and changes in demographics (smaller families).

In 2012, the Estonian data was re-examined at the item level (see Section 4.2.1). The results of that effort are important to the understanding of at least one cause and of an otherwise perplexing difference between Classical Test Theory and Item Response Theory results (see Section 4.9.2).

### 2.2. South Africa

$\Delta IQ = 3.63$  Whites (same group took two different test batteries)

$\Delta IQ = 1.57$  Indians (same group took two different test batteries).

The FE score gain is stronger for the Afrikaans speakers than for the English speakers (te Nijenhuis, Murphy, & van Eeden, 2011).

### 2.3. Gains seen in young children

British children aged 6 and 18 months displayed large developmental gains over the period from 1949 to 1985. When measured on the Griffiths Test, developmental quotients (DQ) gained 2.45 points per decade. Similar studies, using the Bayley Mental Scales (Bayley, 1993) were done by other researchers in the U. S. and Australia and show gains of 2.9 DQ points per decade (Black, Hess, & Berenson-Howard, 2000; Campbell, Siegel, Parr, & Ramey, 1986; Lynn, 2009a; Tasbihsazan, Nettlebeck, & Kirby, 1997). Similarly, Kanaya et al. (2005) reported that elementary school children show FE gains on the WISC that are similar to adult gains on the WAIS. These DQ and IQ gains show a FE that is as large in infants and preschool children as in adults, making education an unlikely explanation for the cause (at least in the data sets examined).

As is already apparent, FE findings in one place do not generalize globally. Cotton et al. (2005) found no FE effect, using the Raven's Colored Progressive Matrices, for

a group of Australian children ages 6–11 from 1975 to 2003; but [Nettelbeck and Wilson \(2004\)](#) found 5 point gain for a range of Australian elementary-grade children from 1981 to 2001.

#### 2.4. Gains in the Raven's Progressive Matrices

The Raven tests have been cited frequently in the FE literature because most samples show particularly large gains on these tests. The Raven and similar tests have shown gains of 18–20 IQ points per generation in many industrialized countries ([Flynn, 1999](#)). Dutch gains were 21 points over 30 years ( $\Delta IQ = 7$ ), while urban Chinese gained 22 points between 1936 and 1986,  $\Delta IQ = 4.4$  ([Neisser, 1998](#)).

[Hiscock \(2007\)](#) found a higher rate of FE gains for the Raven's Progressive Matrices than for the Wechsler and Stanford–Binet tests. He also showed that British Raven scores for birth years from 1877 to 1967 increased steadily, but rolled off over that time span to a possibly flat (no effect) rate for the last 10 year interval.

#### 2.5. Low-end versus high-end gains

As previously mentioned, [Teasdale and Owen](#) found that FE gains for Danish draftees were concentrated in the lower end of the intelligence spectrum, suggesting a cause or causes such as improved nutrition, better health care, or increased education. [Colom, Andre's Pueyo, and Juan-Espinosa \(1998\)](#) noted that FE gains were much greater on the Raven's Standard Progressive Matrices (19.2 points over 28 years,  $\Delta IQ = 6.9$ ) than on the Advanced Progressive Matrices (6.75 points over 28 years,  $\Delta IQ = 2.4$ ). They concluded that the cause of the increases probably had a greater impact in the low and medium segments of the intelligence distribution. In a later study, [Colom et al. \(2005\)](#) also found that gains were more pronounced in the lower range.

[Lynn and Hampson \(1986\)](#) reported a low-end gain that was about double the high-end gain, for a British group over the period 1932 to 1982. Similarly, [Kagiticbasi and Biricik \(2011\)](#) found greater gains in Turkey at the low end, over the period from 1977 to 2010. The differences were particularly large (23 points,  $\Delta IQ = 7$ ) for remote villages. Within urban locations, the lower SES groups also showed more gains (7.4 points,  $\Delta IQ = 2.2$ ) than higher SES groups, but these were less than in the remote villages.

The FE is so specific that for every finding, there seems to be an opposite finding. [Flynn \(1996, 2009\)](#) claimed IQ gains at “every level,” based on his observation that “score variance remains unchanged over time.” His “every level” projection held in a study conducted in La Plata, Argentina, where  $\Delta IQ = 6.3$  and showed no bias towards high or low IQ ranges. Flynn extended this observation as meaning that nutrition is an unlikely explanation, since it would presumably apply more readily to gains seen at the lower end, and not throughout the intelligence spectrum ([Flynn & Rossi-Casé, 2012](#)). [Flynn \(2009\)](#), cited [Sundet, Barlaug, and Torjussen \(2004\)](#)) as an example in which IQ gains were concentrated in the lower half of the IQ spectrum, while height gains were mostly in the upper half, pointing out that this combination is inconsistent with the nutrition argument.

[Colom, Flores-Mendoza, Francisco, and Abad \(2007\)](#) examined data for Brazilian children covering a span of 72 years. They found that the FE gains were greater for urban samples than for rural samples and concluded: “Whatever the causes of the increase, they act more intensively for more intelligent children.”

[Ang, Rodgers, and Wänström \(2010\)](#) computed FE gains from the National Longitudinal Survey of Youth (NLSY) data, which include scores from the Peabody Individual Achievement Test (PAIT); the math portion was deemed to be closest to fluid intelligence. In this instance, the gains were skewed towards more educated and higher income families. Only the PAIT-math showed FE gains, which the authors believe is difficult to explain by a nutrition hypothesis. This study showed no race or sex related differences in FE gains.

#### 2.6. Right tail gains

Only one study examined the FE in a data set that is limited to very high IQ individuals. [Wai and Putallaz \(2011\)](#) examined the huge (1.7 million scores) American data set of 7th grade students who took the SAT and ACT and 5th and 6th grade students who took the EXPLORE test. These tests are given to students who have scored in the top 5% for their grade on a standardized test (composite or subtest), and are part of the Duke Talented Identification Program 7th grade search.

[Flynn \(1996\)](#) argued that the gains were present at all levels, but did not have data specific to the high range that is usually considered as gifted. [Wai and Putallaz](#) found the following generational IQ gains in the top 5%:

5.1	SAT-M
13.5	ACT-M
11.1	EXPLORE-M

The gains were concentrated on math and nonverbal subtests (see previous comments on [Ang et al., 2010](#)).

[Wai and Putallaz](#) also examined SAT-M scores of 500 and above (top 0.5%) and equivalent scores for the ACT, with the following results:

SAT-M 1981–1985, 7.7% at or above 500
2006–2010, 22.7% at or above 500
ACT-M 1990–1995, 17.7% at or above a similar level
2006–2010, 29.3% at or above a similar level.

The obvious conclusion is that either there are a lot more truly bright children in the 2006–2010 set, or the test results are showing a significant score inflation that is not merited. They also used multigroup confirmatory factor analysis to determine whether the data sets were invariant with respect to cohort; they were not. Consequently, it can be concluded that something changed in the test construct from one cohort to the other.

#### 2.7. FE gains but without a change in inspection time

Perhaps the only study to link a biological correlate of intelligence and test scores with the FE was carried out by

Nettelbeck and Wilson (2004) in Australia. In 1981, Wilson conducted a study of school grades 1 through 7, administering the Peabody Picture Vocabulary Test (PPVT) and measured inspection times (IT) for each of the participants. In 2001, the study was replicated with virtually every parameter held constant, other than the students. The study was done in the same school, with the same grade levels, using the same PPVT and the revised PPVT-III. IT was measured with the same Gerbrands tachistoscope, under identical conditions.

The results of the study were that the students in 2001 scored essentially the same on the PPVT-III as did the students in 1981 on the PPVT. The 2001 students scored almost 5 points higher when they took the PPVT ( $\Delta IQ = 2.5$ ). IT measurements were the same to within the error bands. Thus, the FE was shown, but was not accompanied by improvements in IT. I asked Nettelbeck if there were any observable differences in SES or nutrition between the two groups. He said that the area served by the school was stable and that there were no observable differences in such things as nutrition or standard of living.

While IT does not correlate significantly with fluid intelligence (Burns & Nettelbeck, 2003; Burns, Nettelbeck, & Cooper, 1999), it does correlate with nonverbal IQ at about 0.50 (Deary & Stough, 1996; and others) and with Raven's matrices and performance IQ. The finding suggests that FE gains were unrelated to processing speed or other factors that explain the IT to general ability correlations.

### 3. Academic performance down

While IQ test scores have been rising (in some cases soaring), academic performance has done the opposite. As Jensen (1998) pointed out, when he observed that the SAT and subtests of scholastic test items have declined, real world academic performance has done the same.

Adey and Shayer (2006), of King's College London, studied the test scores of 25,000 children across both state and private schools and concluded: "The intelligence of 11-year-olds has fallen by three years' worth in the past two decades. In 1976 a third of boys and a quarter of girls scored highly in the tests overall; by 2004, the figures had plummeted to just 6% of boys and 5% of girls. These children were on average two to three years behind those who were tested in the mid-1990s."

For an assessment of how well U.S. students are doing, this URL leads to a well written, if depressing, description of the state of teaching, education, and students: <http://www.lhup.edu/~dsimanek/decline1.htm>.

### 4. Hypothetical causes

Among the causes that have been proposed to explain the FE are these:

Education	Decreased family size
Increased exposure to testing	Heterosis
Exposure to artificial light	More complex visual environment
Nutrition	Child rearing practices

and the use of Classical Test Theory versus Item Response Theory.

#### 4.1. Education

Since FE gains have been observed in preschool children, education is unlikely to be a cause in all data sets. As previously discussed, FE gains have usually been more pronounced on non-scholastic items, while scholastic subtests have presented lower scores at the same time and within the same tests. Direct measures of academic performance have also shown secular declines while FE gains were evident in IQ tests (Jensen, 1998). Lynn (1998) argued that the Raven tests are being inflated as a result of mathematical education; however, the relationship of simple math to increased education is a questionable factor, especially in the Colored or Standard tests (Carlson & Jensen, 1980).

Rönnlund and Nilsson (2008, 2009) examined data from the Betula prospective cohort study. This Swedish data set consists of four age-matched samples (35–80 years;  $N = 2996$ ) tested on the same battery of memory tasks. Data was taken in 1989, 1995, 1999, and 2004. A FE was found at  $\Delta IQ = 1.5$  (relatively low, relative to other nations). FE gains in fluid and crystallized intelligence were approximately equal. Years of education, height (interpreted as a marker for nutrition), and sibsize were used as markers; together they accounted for over 94% of the time-related differences in cognitive performance. But education was a much stronger predictor than the other two items. The authors wrote: "The fact that education emerged as the strongest predictor across all cognitive measures enforces the conclusion that education may exert influence on time-related patterns on (broad) fluid (visuospatial ability, episodic memory) as well as crystallized/semantic aspects of cognition."

#### 4.2. Increased exposure to testing

There is little doubt that testing frequency has increased over the past years. Tuddenham listed it as one possible explanation for the secular gains he found between WW1 and WW2 cohorts. There are two mechanisms that have been proposed. Brand (1996) suggested that the use of timed tests has caused students to work faster by guessing more frequently on multiple choice tests. This largely ignored hypothesis has recently been supported by item level data (Must & Must, 2012). This finding explains other observations (lack of  $g$  loading in some studies and inconsistency between scoring methodologies) but does not cover all aspects of this category of causation. For example, FE gains are seen on tests that are untimed and on tests that do not use multiple choice.

Jensen (1998, p. 327) mentioned "increasing test wiseness from more frequent use of tests." His point was that frequent testing may have the same sort of impact on test scores as the increase associated with test–retest. This is the same process that is associated with learning and shows up in situations where test training has been used (as is common with the SAT). When this happens, the test  $g$  loading decreases and its  $s$  loading (specificity) increases.

Both Brand's and Jensen's ideas would presumably cause test scores to increase without showing gains on  $g$ . As will be seen later, numerous studies, but not all, have shown that FE gains that are not  $g$  loaded. Flynn (2009) agreed with Jensen's comment (above), but only for the early years



of testing: “The twentieth century saw us go from subjects who had never taken a standardized test to people bombarded by them, and, undoubtedly, a small portion of gains in the first half of the century was due to growing test sophistication. Since 1947, its role has been relatively modest.”

#### 4.2.1. Estonian data supports Brand's hypothesis

Brand (1996) wrote: “The correct strategy for testees is: ‘When in doubt, guess.’” This hypothesis has been occasionally noted in the literature, but seldom described as a likely and significant driver of FE gains.

Item level data was preserved for the Estonian National Intelligence Test, from 1933/36 and 2006. These data show a change in test taking strategy that is best described as increased guessing (Must & Must, 2012). The numbers of correct answers increased (SD .79), but that increase was accompanied by an increase in incorrect answers (SD .15). The number of missing answers decreased. Scores were not penalized by wrong answers, but were boosted by correct answers. The Estonian data showed relatively little guessing effect for comparisons and other simple tasks, but had a large presence on time-pressured and mentally taxing tasks (math). In the 1934–36 tests the item level data do not suggest the guessing strategy that is apparent in the 2006 tests. It should be noted that these same data show FE gains in excess of those that can be attributed to a guessing strategy.

### 4.3. Nutrition and medical care

Both nutrition and medical care have improved over the past century and have been accompanied by a large number of gains that appear to be caused by these improvements: increased mean height, increased head size, faster growth, earlier maturation, etc. Lynn (2009a) argues that gains in developmental quotients (DQs – hold up head, sit up, stand, walk, jump, etc.) are indicators of gains in IQ. DQs have gained 3.7 points per decade, while IQ gains of 3.9 points per decade have been seen in preschool children (age 4–6). Using the Griffiths Test, British children at age 6 months showed an average DQ gain of 2.8 points per decade and children, age 18 months, showed an average gain of 2.1 points per decade. Flynn (1984b) and Bocerean, Fischer, and Flieller (2003) have reported IQ gains that are similar to the DQ gains (Hanson, Smith, & Hume, 1985) for preschool children.

Lynn (2009a,b) cites various studies that show poor nutrition in the early part of the 20th century in the U.S. and Western Europe. Those indications of poor nutrition disappeared over the course of that century. Three nutrients that are known to be related to the development of intelligence are iron, folate, and iodine. Lynn (2009a) presented references showing insufficient intake of these in various countries in the early part of the 20th century. Liu et al. (2012), pointed to improvements in standard of living, nutrition, and education as possible causes for the gains in China. The studies that have shown greater FE gains in the lower part of the IQ distribution are consistent with the nutrition argument.

#### 4.3.1. Birth weights

One factor influencing birth weight is pre-natal nutrition. Birth weight correlates positively with IQ and with DQs. Brazelton, Tronik, Lechtig, Lasky, and Klein (1977) reported that when birth weights reached 3500 g, infants were advanced by approximately 15 DQ points at age 28 days (compared with lower birth weight babies). Low birth weights show the opposite; Drillien (1969) reported DQ score depressions of 12 points for infants with birth weights under 2000 g, compared to those with birth weights over 2500 g (ages 6 months through 2 years). Various other studies have reported similar findings. In general, improved pre-natal nutrition increases birth weights and head size [birth weight is correlated with head size at  $r = 0.75$  (Broman, Nichols, & Kennedy, 1975)]. It is head size that is directly linked to higher cognitive performance.

#### 4.3.2. Height

Lynn (2009a) attributes the change in height and in DQs as being caused by nutritional improvements. Both measures increased by about one standard deviation (SD) over 50 years. Flynn (2009) countered that gains in height have not happened at the same times as gains in IQ. This argument seems to imply a degree of data tracking, with respect to time, that is not necessary for the argument to hold (Lynn, 2009a). Height and intelligence gains for Norwegian conscripts were reported by Sundet et al. (2004) continuing until the late 1980s, when height gains ended. For the period from 1969 to 2002, the height gains were more pronounced in the upper half of the distribution, while intelligence gains were greater in the lower half.

#### 4.3.3. Head size

Lynn (2009a) cited numerous sources that have reported head size increases of about one standard deviation over the past 50-plus years. In Britain, the head circumference of 1 year olds has increased by approximately 1.5 cm from 1930 to 1985 (Cole, 1994). Head circumference, DQs, IQs, and height, over that time span, have all shown gains of about 1 SD. Head size is an approximate measure of brain size; the two correlate at  $r = 0.8$  (Brandt, 1978).

Jensen (1998) found that head size is mostly correlated with  $g$  (as opposed to group factors) and notes that the reason for the correlation is that head size is a proxy for brain size. When measured with MRI, the correlation between brain size and IQ is about 0.40 (Rushton & Ankney, 1996). Larger brain size means more neurons and is logically consistent with the correlations between head and brain measurements versus IQ.

The correlation between brain volume and IQ is presumably due to the larger number of neurons in larger brains (Rushton & Ankney, 1996), although Miller (1994) has suggested that it may be due to higher levels of myelination in larger brains. In any case, increases in brain size should be direct contributors to higher intelligence (Miller & Penke, 2007).

#### 4.3.4. Not nutrition

- Neisser (1998) pointed out that studies of nutrition have shown that neither vitamins nor supplements have had any impact on intelligence.

- Nutrition is unlikely to have declined over the past 20 years in those countries that have a negative FE; height did not decline.
- Contrary to the intelligence gains seen in Norway, height gains from 1969 to 2002 were mostly in the upper half of the intelligence range (Sundet et al., 2004).
- With the exception of Spain, Denmark, and Norway, gains have not been frequently concentrated in the bottom half of the distribution. Flynn and Rossi-Casé (2012) argued that for all other cases, the nutrition argument is not viable.
- Mingroni (2007) argued that all postnatal environmental factors are implausible because of the high consistency of heritability estimates.
- Mingroni (2007) also contended that heterosis is a better explanation for increases in height than are nutritional and health care considerations.

#### 4.4. Exposure to artificial light

This hypothesis is not seen often in the literature and might have been omitted in this review, except that it did not come from a weak source, but was one of the items listed by Jensen in *The g Factor*. The idea is based on the response of the pineal gland in animals to artificial light. The pineal gland appears to play a major role in sexual development, hibernation, metabolism, and seasonal breeding. Artificial light is used by poultry farmers to stimulate growth and increase their output.

There does not seem to be any data available for whether this effect happens in humans, but the speculation is that it might. There has been an obvious increase in the use of electric lighting by humans over much of the time that the FE has been observed. Besides lighting, people have been increasingly exposed to artificial light from television and computer screens, even during early childhood.

#### 4.5. Decreasing family size

It has been known for some time that the mean IQ of families decreases as family size increases. There are two factors that contribute (presumably independently) to this effect:

- Maternal IQ correlates negatively with fertility. This is the underlying factor behind Richard Lynn's papers and book relating to global dysgenics and has been shown for numerous data sets from various countries (Lynn, 1996; Lynn & Harvey, 2008). Low IQ people statistically have more children than high IQ people. The high heritability of intelligence, therefore, is a source of dysgenic pressure. If there is a decrease in average family size (not limited to the upper end), the reduced numbers of low IQ children should produce a net increase in the mean, which would show up as a FE gain.
- Dating as far back as Sir Francis Galton, it was believed that IQ declined as a function of birth order. That belief was disputed by Rodgers, Cleveland, van den Oord, and Rowe (2000) after they examined the American NLSY data and did not find a birth order effect. This argument seemed strong and held until Bjerkedal, Kristensen, Skjeret, and Brevik (2007)

published a study based on a very large data set of Norwegian conscripts, which showed the birth order effect in Norway. The mechanism of the effect has not been resolved. Hypotheses that have been advanced include prenatal gestational factors and social factors. The former seem more consistent with the general finding that social factors have little, if any effect on intelligence. Causation of the birth order effect does not matter with respect to the FE. If family size is declining in various groups, there must be a positive contribution to mean IQ due to fewer low IQ children being born.

#### 4.6. Heterosis

Mingroni (2004, 2007) suggested that since the effects of the environment on intelligence are so small (Loehlin, Horn, & Willerman, 1989; Scarr & Weinberg, 1978), the possibility of a genetic effect should be investigated. If environmental factors were significant, between-family variance would cause MZA twins (identical, reared apart) to be less alike and siblings to be more alike.

Besides IQ, there have been secular trends in height, growth rate, myopia, asthma, autism, ADHD, and head circumference. It may, therefore, seem reasonable to argue that there is a global change that is affecting some or all of these factors (possibly consistent with Lynn's nutrition hypothesis). If selective breeding was involved, in order to produce the magnitudes seen in the FE, breeding would have to be restricted to only those people in the upper half of the IQ distribution (Jensen 1998, p. 327). As previously discussed, it is the bottom half that has the higher fertility.

Lynn (2009a) argued that heterosis is unlikely for three reasons:

- 1 There was little immigration in Europe before 1950 (the FE was present before that date).
- 2 The FE for IQs and DQs is just as large in Europe as in other places.
- 3 Studies of heterosis have shown little positive effect on IQ.

Woodley (2011) also concluded that heterosis is an unlikely cause because the FE gains are seen on the least *g* loaded components of intelligence tests [Colom, Juan-Espinosa, and Garcia (2001) reported opposite findings for Spanish standardizations of the DAT.].

Perhaps the most important consideration in determining whether there is a heterosis effect was pointed out by Mingroni: If the FE is found within-families, the cause is not genetic. Sundet, Eriksenb, Borren, and Tamsb (2010) found that the FE operates within sibships. Unless this finding cannot be extended beyond Norway, the heterosis hypothesis does not look viable.

Mingroni (2007) argued in favor of a heterosis explanation from the perspective of real gains on intelligence and did not address situations, such as increased exposure to testing (Section 4.2), that show a FE, but which are inherently not Jensen effects. He also argued that increases in height were better explained by heterosis than by nutrition, but did not address that at least some of the height gains are related to leg length and are best explained by sexual selection (Jensen, 1998, p. 331).

#### 4.7. Enriched visual environment

Greenfield (1998) and others suggested that the FE gains are caused by the ever increasing shift from verbal communication to visual and interactive media. This is seen globally in the increased presence of movies, television, photography, video games, computers, puzzles, mazes, exploded views, etc. Advertising has become ubiquitous and is saturated with images, graphs, charts, and rapid sequence visuals.

The mechanism for this hypothesis is that the shift towards visual representations removes some of the novelty from tests, especially in the culture reduced tests that have shown about double the FE gains as found in other tests. This is particularly convincing for tests such as the Raven which presents abstract figures in a matrix. Several decades ago these figures may have been more baffling than they are today.

#### 4.8. Child rearing practices

The FE has been seen throughout the world, in both developed and undeveloped countries where child rearing practices vary greatly. It is unlikely that this hypothesis is a significant factor, not only because of the cultural variation in child rearing practices, but also because the shared environment has essentially no impact on adult intelligence (per prior discussion). To some extent, this category overlaps the increased visual environment and education. In that regard, it may contribute to the FE in some instances.

#### 4.9. Methodological and test construct issues

As previously mentioned, ceiling effects can distort FE measurements. Other methodological issues have been found, but not fully resolved.

##### 4.9.1. Is the FE invariant?

When researchers have tested for invariance, they have found that the data sets they were examining were not invariant (Must et al., 2009; Wai & Putallaz, 2011). Wicherts et al. (2004) did a study of five data sets to test for invariance. These included the Must et al. and Teasdale & Owen studies. Multigroup confirmatory factor analyses of these data sets showed that they were not invariant, meaning that FE gains were not gains on the latent variables that the tests were supposed to measure. Besides providing insight as to the nature of the FE gains, the rejection of factorial invariance demonstrates that subtest score interpretations are necessarily different over time.

Flynn (2009) pointed out that cultural changes over time cause some test items to become easier because they have lost their novelty. Some words that were previously not common become more common because usage has changed. He gives several examples of this, including his frequently used example: “What do dogs and rabbits have in common?” He says that past generations would more likely focus on the use of dogs to hunt rabbits, while later generations would immediately identify that they are both mammals. This example of differential item functioning is probably responsible for at least some subtest score increases, especially in tests of similarities and vocabulary. Periodic test revisions should remove these non-g gains.

##### 4.9.2. Classical Test Theory versus Item Response Theory

Beaujean and Osterlind (2008) did an analysis that is related to the Wicherts et al. analysis of invariance, which examines the underlying nature of the test itself. Most studies in the literature are based on Classical Test Theory (CTT) and present results which are not based on item level analysis. This practice hides some of the information that could be extracted from a data set. Test scores are given, but the latent constructs they are designed to measure cannot be examined. Item Response Theory (IRT), on the other hand, allows the researcher to examine the changes in underlying latent ability. Thus, CTT can show differences in scores, even when there is no change in the latent variable. An increase may be due to a general gain in real intelligence, or a decrease in the levels of difficulty of test items.

Despite its relatively infrequent use, IRT is generally considered to be the better methodology. It is particularly useful in FE studies because it reveals changes in item properties between two groups measured at different times. CTT requires groups that are being compared to have similar ability distributions, but this is not a requirement when IRT is used. In IRT, the item parameters do not depend on the ability level of the testees.

Results using CCT and IRT to measure FE gains in the American NLSY data:

<i>Peabody Picture Vocabulary Test-Revised (PPVT-R)</i>	
CCT	0.44 points per year
IRT	0.06 points per year
<i>Peabody Individual Achievement Test-Math (PIAT-M)</i>	
CCT	0.27 points per year
IRT	0.13 points per year

The results show that the FE essentially vanishes for the PPVT-R when IRT is used. The PIAT-M gains are cut to half using IRT. Ergo, the FE gains are a function of the methodology, leading to the concern that much of the literature has reported findings that might be quite different if IRT had been used.

Now that an item level study has been reported for the Estonian data (see Section 4.2.1), it is apparent that some of the score gains were due to increased guessing on the most complex subtests. Shiu, Beaujean, Must, te Nijenhuis, and Must (2012) reported effect sizes for the FE gains in this data set. All subtests, except computations, showed gains; the largest gain was in analogies. The research group concluded that there was some real increase in abilities (beyond the guessing related gains previously discussed).

## 5. Real or hollow gains?

When David Wechsler studied his WAIS, he gave the old 1953 version and the new revised 1978 version (WAIS-R) to the same group. That group averaged 103.8 on the new version and 111.3 on the old version yielding  $\Delta IQ = 3$  (Neisser, 1997).

If children of 1997 took the 1932 Stanford-Binet, 1/4 would score above IQ 130 (an increase of  $10\times$ ). If children in 1932 took the 1997 test, the mean would be about 80! 1/4 would be “deficient” (Neisser, 1997).

Vroon made a similar observation about Dutch men: When scored against 1982 norms, men in 1952 would have had a mean IQ of 79 (Neisser, 1998).

Flynn initially questioned the reality that intelligence has increased:

“Has the average person in The Netherlands ever been near mental retardation?” “Does it make sense to assume that at one time almost 40% of Dutch men lacked the capacity to understand soccer, their most favored national sport?” He noted that there are not more gifted Dutch school children now and that patented inventions have shown a sharp decline. The U.S. mean in 1918 would have been 75, if scored against today’s norms. If the score gains were real intelligence gains, real-life consequences would be conspicuous (Neisser, 1998). In discussing paradoxes related to the secular gains, Flynn (2009) wrote: “How can people get more intelligent and have no larger vocabularies, no larger stores of general information, no greater ability to solve arithmetical problems? ... Why do we not have to make allowances for the limitations of our parents?”

### 5.1. Is the Flynn effect a Jensen effect?

[A Jensen effect is one that loads on *g*. It was named by Rushton.]

- Colom et al. (2001) Paper title: The secular increase in test scores is a “Jensen effect.”
- Must et al. (2003) Paper title: The secular rise in IQs: In Estonia, the Flynn effect is not a Jensen effect.
- Rushton and Jensen (2010): “The Flynn effect is not a Jensen effect (because it does not occur on *g*).”

#### 5.1.1. Not a Jensen effect

In a meta-analysis of 64 test–retest studies using IQ batteries (total  $N = 26,990$ ), te Nijenhuis, van Vianen, and van der Flier (2007) found a correlation between *g* loadings and score gains of  $-1.00$ . A similar finding was reported for a different meta-analysis by van Bloois, Geutjes, te Nijenhuis, and de Pater (2009). Must et al. (2003) found (in Estonia) a correlation of  $-0.40$  between *g* and FE gains. These all show that the gains were not on *g* and were, therefore, hollow. The discussion in Section 4.2.1 shows that at least part of the Estonian gains were the result of an increased tendency to guess.

Rushton and Jensen (2010) showed that heritabilities calculated from twins also correlate with the *g* loadings,  $r = 0.99$ ,  $P < 0.001$  (for the estimated true correlation), providing biological evidence for a genetic *g*. The importance of this is that if the FE is being driven by environmental factors, it is unlikely that the gains would load on *g*. If the cause is genetic (as in the Mingroni hypothesis), the gains should show a Jensen effect.

They also pointed out that *g* loadings and inbreeding depression scores on the 11 subtests of the WISC correlate significantly positively with racial differences and significantly negatively (or not at all) with the secular gains. This is further evidence that the FE is caused by environmental factors.

Perhaps the strongest argument that the FE does not load on *g* came from Rushton (1999). He used principal components analysis to show the independence of the FE from known genetic effects.

- The IQ gains on the WISC-R and WISC-III form a cluster. This means that the secular trend is a reliable phenomenon.
- This cluster is independent of the cluster formed by racial differences, inbreeding depression scores (purely genetic), and *g* factor loadings (largely genetic). The secular increase is, therefore, unrelated to *g* and other heritable measures.

Must et al. used the Method of Correlated Vectors (see Jensen, 1998) to test the FE gains for *g* loading. Rank order correlations between the various subtests and the rank of those subtests on the *g* factor were negative and nonsignificant:  $r = -.40$  (one-tailed  $P = .13$ ). Subtests with the lowest *g* loadings showed the greatest FE gains. The authors concluded: “In Estonia, the Flynn effect is not a Jensen effect.”

#### 5.1.2. Yes, it is a Jensen effect

Colom et al. (2001) examined two successive Spanish standardizations of the Differential Aptitude Test (DAT) battery and found gains on *g*,  $r = .78$ ;  $P < .05$ . Colom: “Not a ‘Jensen effect’ is true for crystallized tests but not for fluid tests.” Using the DAT, Colom et al. showed that subtest gains increased as their rank order of *g* loading increased [the subtests in the DAT are (in order of increasing *g* loading) numerical ability, verbal reasoning, mechanical reasoning, abstract reasoning, and spatial relations.].

### 5.2. Predictive bias

Jensen (1998, p. 331) stated that the definitive test of whether FE gains are hollow or not is to apply the predictive bias test. This means that two points in time would be compared on the basis of an external criterion (real world measurement, such as school grades). If the gains are hollow, the later time point would show underprediction, relative to the earlier time. This assumes that the later test has not been renormed. In actual practice tests are periodically renormed so that the mean remains at 100. The result of this recentering is that the tests maintain their predictive validity, indicating that the FE gains are indeed hollow.

## 6. Which explanations work?

Most of the mechanisms that have been proposed as causes of the FE are plausible under some circumstances. Even when one is ruled out by a specific study, it may apply elsewhere. As has been shown in the foregoing material, the most consistent aspect of the FE is that it is inconsistent from one time or place to another. Sometimes the gains have been mostly in abstract reasoning (as in the U.S.), but elsewhere the gains have been strongly tilted towards scholastic subtests (Estonia). Gains have been strong, weak, flat, or have reversed, even within the same country when measured at different times — Norway and Denmark (Sundet et al., 2004; Teasdale & Owen, 2008).

Finally, there are the issues of non-invariance and of methodological inconsistency when IRT is used instead of



CTT. The instances in which confirmatory factor analysis has failed to show invariance (every case so far) tell us that the meaning of IQ tests is not constant over time. The reduction in FE magnitude (to near zero in some cases) when IRT is applied suggests that the test vehicle is contributing 50 to 100% of the gains and that those gains are methodological artifacts and carry no  $g$  loading. For example, the FE gains due to guessing (Estonia) were not resolved by CTT because the successful strategy was not apparent at the subtest level.

### 6.1. Real or hollow?

Most of the tests for  $g$  loading have shown little or no  $g$  saturation. The majority of researchers who have addressed the issue have argued that the gains are hollow, with the exception of Lynn and Colom, both of whom have made strong arguments that there is at least some genuine gain in intelligence. This inconsistency may be due in part to different data sets and may be due in part to CTT methods. It is likely that most of the FE gains that have been reported are hollow. If this were not true, renorming would cause predictive validity to change, but there are no reports that this has happened.

## 7. Can the Flynn effect be modeled?

Most studies of the FE have attempted to apply a single explanation, such as heterosis, or a narrow category of causation, such as nutrition/health care. This overview, however, strongly suggests that multiple causes are acting, and that the mix of causes varies over time and from one place to another. Flynn and Rossi-Casé (2012) agree: “Even in developed nations, the notion that the Flynn effect will have identical causes should be banished from the literature.”

A quantitative model of causation is beyond present understanding, but a qualitative model can be constructed, such that the most likely active components can be identified. Two approaches to this follow.

### 7.1. A life history model

Woodley (2012) presented a model in which a large number of FE causes (as discussed here) are assumed to vary as a group. His model assumes that the FE gains are unrelated to  $g$  and are the result of a shift in life history from fast to slow. A fast life history is taken to be the set of tradeoffs that are associated with relatively high fertility and lower parental investment in offspring, as described by Rushton (1985) in his Differential K Theory; slow life history is the opposite (lower fertility and more parental investment). Woodley describes his model as a cognitive differentiation–integration effort (CD–IE) hypothesis.

- Cognitive integration effort (CIE) – a strengthening of the manifold via the investment of bioenergetic resources – fast life history.
- Cognitive differentiation effort (CDE) – a weakening of the manifold via the unequal investment of resources into individual abilities – slow life history.

If it happens that a given population is moving from a fast towards a slow life history, multiple environmental factors can be expected to move in the direction that would cause a secular rise in test scores: fertility, education, pathogen stress, and nutrition.

### 7.2. Independent Drivers model

The Woodley model, described above, focuses on a latent variable, such that variations in that variable contribute to the FE by means of the causes that are assumed to increase or decrease together. An alternative model assumes that the various FE drivers act independently, may combine in any combination, and may include negative driver components. The causes that are present in a given data set over an observation period are difficult to quantify, but can be estimated on a limited scale, such as high, medium, and low, with the expectation that their contributions to FE gains will be larger or smaller, depending on the strength of the driver.

Each driver is assumed to exert a FE influence as a function of how much contribution potential remains in association with that driver. For example, the reduction in family size is likely to initially contribute more to a study group that has had high fertility and is moving in the direction of smaller families. As the process continues, diminishing FE gains will be seen as the maximum total effect is used up. The path may appear to be somewhat linear over a short time period, but it must approach an asymptote. The gain for any given driver should follow a relationship that is similar to

$$FEG_i = FEM_i(t)/(t + k_i)$$

where  $FEG_i$  is the FE gain due to driver  $i$ ;  $FEM_i$  is the maximum FE gain that can be contributed by driver  $i$ ;  $t$  is the time in years; and  $k_i$  is a constant for driver  $i$ . Multiple drivers would be additive, but each will have its own maximum contribution and constant.

The shape illustrated in Fig. 1 is consistent with the gains (general shape) shown by the Raven's Progressive Matrices in Britain (Hiscock, 2007).

#### 7.2.1. Reversals

Reversals may occur either as the sum of positive drivers decreases to less than the sum of negative drivers, or the positive drivers reverse direction. A lack of FE push might result in a reversal due to an existing negative cause, such as an underlying dysgenic trend or the decline in educational participation. The net FE gain (or loss) may contain negative factors that are not evident in the data, because the result is a positive FE. Thus, the positive drivers need only reach saturation for a reversal to appear (assuming the presence of one or more negative drivers).

It is possible that some of the drivers that have been discussed could reverse direction and directly cause a FE decline. For example, nutritional factors may change and become negative due to the introduction of harmful chemicals into diets or the living environment; health care standards could deteriorate; family sizes could reverse direction, at least for a segment of a population.

## 7.2.2. FE Drivers

Group and environmental characteristics over the time period $\Delta T$	FE driver
More school years completed Qualitatively improved education Higher scores on scholastic tests Score gains seen in preschool children	Education  Not education, but possibly nutrition, health care, etc. Increased exposure to testing
More testing in primary and secondary schools Increased use of tests for college level selection Recent electrification, as might be seen in remote areas Increased availability of television Growth of personal computers in homes and schools	Exposure to artificial light
Increased pediatric care Diet improvements of critical nutrients Mean increases in height, head size, or birth weight Accelerated childhood development	Nutrition and health care
Lower fertility for low SES levels Increased availability of television Growth of personal computers in homes and schools Increased visual complexity of school textbooks Advertising growth, accompanied by charts, symbols, etc. Measured increase in mean $g$	Decreased family size More complex visual environment  Nutrition and health care Decreased family size
Change in breeding pattern from isolated groups to breeding among groups, not accompanied by within-family FE	Heterosis

For a given data set, the presence of items from the first column implies a cause from the second column. For example, [Must and Must \(2009\)](#) reported a height increase (in Estonia)

of 2.9 SD over approximately 2 centuries. At the beginning of the 20th century, the diet was primarily bread and herring. From 1925 to 1958 there was a shift from vegetarian foods to meats. This pattern signals that the nutritional FE driver was active during and after the dietary change. FE gains were seen in scholastic performance and reasoning, suggesting that education was also a factor. The general increase in prosperity of the country may also signal matches for other changes (first column), such as decreased family size.

In some situations, the Independent Drivers model could reduce to the Woodley model, but in situations where the effect can only be linked to one or two drivers, this model is accommodating. In any situation where a gain in  $g$  is seen, the Woodley model would not apply, but this model identifies nutrition, health care, and heterosis as possible  $g$  loaded drivers.

## 8. Summary

- The FE exists between birth cohorts.
- It has been found within sibships.
- It sometimes appears early in life (before school age).
- There are presumably multiple causes.
- The gains are often hollow (not Jensen effects) but some gains appear to be on  $g$ .
- There are methodological issues to be resolved which may be a cause of some of the gains.
- The FE is not invariant over time.

## 9. Recommendations

Despite the huge mass of papers, the FE remains enigmatic. Part of the problem is the complication of what strongly appears to be varying combinations of multiple drivers; individual studies cannot be consistently compared. But the concern that deserves particular attention is that methodological issues appear to be confounded with real world causes. Perhaps ways can be found to examine more data sets with IRT. It would be

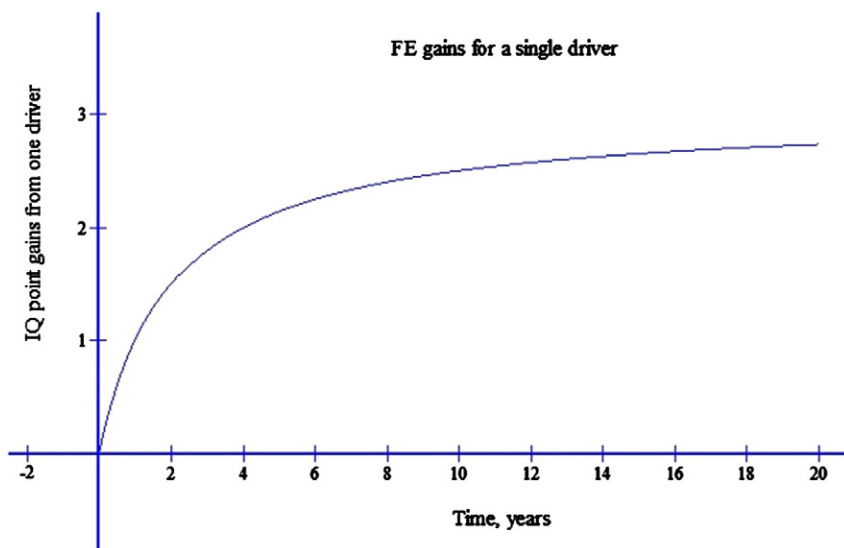


Fig. 1. Flynn effect gains for a single driver. In the illustration the maximum contribution for the driver is shown as 3 IQ points and the value of  $k$  is set at 2.

very helpful to know how much of the various FE gains are the result of CTT methodology. The findings of non-invariance presumably mean that some FE gains are attributable to test revisions and to cultural shifts. A better grasp of the categories of test items that are causing non-invariance may enable test designers to reduce or eliminate these test-specific items.

Some direct connections between environmental conditions and the FE have been identified, such as those in Estonia (dietary changes, family size reductions, and educational improvements). These point to causes for a single country, but cannot be generalized. Future researchers should be encouraged to examine national data sets from health and social service agencies to identify sharp changes that correspond to FE rate changes. Some of this has already been done by Lynn, but there may be additional factors that have not yet surfaced. In the U.S. the National Institute of Health and the Food and Drug Administration are probable data sources. Other environmental factors that might be worth examining for coincidence with FE rate changes: the introduction of radio, television, computers, the Internet, and cell phones, etc. Educational policies and numbers of graduates might be considered as well, despite declines in academic performance, there may still be FE drivers associated with formal or informal education.

Finally, it would be helpful to perform studies of biological parameters that relate to intelligence. There is the IT study by Nettelbeck & Wilson, but little else in this category. The question to answer is whether other biological measurements (RT, brain pH, nerve conduction velocity, pitch discrimination, EEG latencies, glucose uptake rates, etc.) remain stable over decades, or do they vary in the direction that would be predicted by an increase in intelligence?

## Acknowledgment

I would like to thank James Thompson for his constructive comments on this manuscript.

## References

- Adey, P., & Shayer, M. (2006). Cited by Guardian Co Uk, (available at <http://education.guardian.co.uk/schools/story/0,1693061,00.html>).
- Ang, S., Rodgers, J., & Wänström, L. (2010). The Flynn effect within subgroups in the U.S.: Gender, race, income, education, and urbanization differences in the NLSY-Children data. *Intelligence*, 38–4, 367–384.
- Bayley, N. (1993). *Bayley scales of infant development*. San Antonio, TX: Psychological Corporation.
- Beaujean, A. A., & Osterlind, S. J. (2008). Using item response theory to assess the Flynn effect in the National Longitudinal Study of Youth 79 Children and Young Adults Data. *Intelligence*, 36(5), 455–463.
- Bjerkedal, T., Kristensen, P., Skjeret, G. A., & Brevik, J. I. (2007). Intelligence test scores and birth order among young Norwegian men (conscripts) analyzed within and between families. *Intelligence*, 35–5, 503–514.
- Black, M. M., Hess, C. R., & Berenson-Howard, J. (2000). Toddlers from low-income families have below normal mental, motor and behavior scores on the revised Bayley Scales. *Journal of Applied Developmental Psychology*, 21, 655–666.
- Bocerean, C., Fischer, J. -P., & Flieller, A. (2003). Long term comparison (1921–2001) of numerical knowledge in 3 to five and a half year old children. *European Journal of Psychology of Education*, 18, 405–424.
- Brand, C. (1996). *The g factor: General intelligence and its implications*. Chichester, England: Wiley.
- Brandt, I. (1978). Growth dynamics of low-birth weight infants with emphasis on the perinatal period. In F. Falkner, & J. M. Tanner (Eds.), *Human growth, Vol. 2*. (pp. 557–617) New York: Plenum.
- Brazelton, T. B., Tronik, E., Lechtig, A., Lasky, R. E., & Klein, R. E. (1977). The behavior of nutritionally deprived Guatemalan infants. *Developmental Medicine and Child Neurology*, 19, 364–372.
- Broman, S. H., Nichols, P. L., & Kennedy, W. A. (1975). *Preschool IQ: Prenatal and developmental correlates*. Hillsdale, NJ: Wiley.
- Burns, N., & Nettelbeck, T. (2003). Inspection time in the structure of cognitive abilities: Where does IT fit? *Intelligence*, 31, 237–255.
- Burns, N. R., Nettelbeck, T., & Cooper, C. J. (1999). Inspection time correlates with general speed of processing but not with fluid ability. *Intelligence*, 27, 37–44.
- Campbell, S. K., Siegel, E., Parr, C. A., & Ramey, C. T. (1986). Evidence for the need to renorm the Bayley Scales of Infant Development based on the performance of a population-based sample of 12 month old infants. *Topics in Early Childhood Education*, 6, 83–96.
- Carlson, J. S., & Jensen, C. M. (1980). The factorial structure of the Raven Coloured Progressive Matrices Test: A reanalysis. *Educational and Psychological Measurement*, 40, 1111–1116.
- Cole, T. J. (1994). Growth charts for both cross-sectional and longitudinal data. *Statistics in Medicine*, 13, 2477–2492.
- Colom, R., Andre's Pueyo, A., & Juan-Espinosa, M. (1998). Generational IQ gains: Spanish data. *Personality and Individual Differences*, 25(5), 927–935.
- Colom, R., Flores-Mendoza, C. E., Francisco, J., & 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., Juan-Espinosa, M., & Garcia, L. F. (2001). The secular increase in test scores is a "Jensen effect.". *Personality and Individual Differences*, 30, 553–559.
- Colom, R., Lluís-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*, 33, 83–91.
- Cotton, S. M., Kiely, P. M., Crewther, D. P., Thomson, B., Laycock, R., & Crewther, S. G. (2005). A normative and reliability study for the Raven's Colored Progressive Matrices for primary school aged children in Australia. *Personality and Individual Differences*, 39, 647–660.
- Deary, I. J., & Stough, C. (1996). Inspection time and intelligence: Achievements, prospects and problems. *American Psychologist*, 51, 599–608.
- Drillien, C. M. (1969). School disposal and performance for children of different birthweight born 1953–1960. *Archives of Diseases in Childhood*, 44, 562–570.
- Flynn, J. R. (1984a). IQ gains and the Binet decrements. *Journal of Educational Measurement*, 21, 283–290.
- Flynn, J. R. (1984b). 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. (1996). What environmental factors affect intelligence: The relevance of IQ gains over time. In D. Dettnerman (Ed.), *Current topics in human intelligence, vol. 5: The environment*. Norwood, NJ: Ablex.
- Flynn, J. R. (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist*, 54(1), 5–20.
- Flynn, J. R. (2009). *What is intelligence? Beyond the Flynn effect*. Cambridge: Cambridge University Press.
- Flynn, J. R., & Rossi-Casé, L. (2012). IQ gains in Argentina between 1964 and 1998. *Intelligence*, 40, 145–150.
- Greenfield, P. M. (1998). The cultural evolution of IQ. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 81–123). Washington, DC: American Psychological Association.
- Hanson, R., Smith, J. A., & Hume, W. (1985). Achievements of infants on items of the Griffiths scales: 1980 compared with 1950. *Child: Care, Health and Development*, 11, 91–104.
- Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.
- Hiscock, M. (2007). The Flynn effect and its relevance to neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, 29(5), 514–529.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Kagitcibasi, C., & Biricik, D. (2011). Generational gains on the draw-a-person IQ scores: A three-decade comparison from Turkey. *Intelligence*, 39, 351–356.
- Kanaya, T., Ceci, S. J., & Scullin, M. H. (2005). Age differences within secular IQ trends: An individual growth modeling approach. *Intelligence*, 33, 613–621.
- Liu, J., Yang, H., Li, L., Chen, T., & Lynn, R. (2012). An increase of intelligence measured by the WPPSI in China, 1984–2006. *Intelligence*, 40, 139–144.
- Loehlin, J. C., Horn, J. M., & Willerman, L. (1989). Modeling IQ change: Evidence from the Texas Adoption Project. *Child Development*, 60, 993–1004.
- Lynn, R. (1982). IQ in Japan and the United States shows a growing disparity. *Nature*, 297, 222–223.
- Lynn, R. (1996). *Dysgenics: Genetic deterioration in modern populations*. Praeger Publishers.
- Lynn, R. (1998). In support of nutrition theory. In U. Neisser (Ed.), *The rising curve*. Washington, DC: American Psychological Association.

- Lynn, R. (2009a). What has caused the Flynn effect? Secular increases in the development quotients of infants. *Intelligence*, 37(2009a), 16–24.
- Lynn, R. (2009b). Fluid intelligence but not vocabulary has increased in Britain, 1979–2008. *Intelligence*, 37, 249–255.
- Lynn, R., & Hampson, S. (1986). The rise of national intelligence. Evidence from Britain, Japan and the United States. *Personality and Individual Differences*, 7, 23–32.
- Lynn, R., & Harvey, J. (2008). The decline of the world's IQ. *Intelligence*, 36, 112–120.
- Miller, E. M. (1994). Intelligence and brain myelination: A hypothesis. *Personality and Individual Differences*, 17, 803–832.
- Miller, G. F., & Penke, L. (2007). The evolution of human intelligence and the coefficient of additive genetic variance in human brain size. *Intelligence*, 35, 97–114.
- Mingroni, M. A. (2004). The secular rise in IQ: Giving heterosis a closer look. *Intelligence*, 32, 65–83.
- Mingroni, M. A. (2007). Resolving the IQ paradox: Heterosis as a cause of the Flynn effect and other trends. *Psychological Review*, 114, 806–829.
- Must, O., & Must, A. (December 19). The biological correlates of the Flynn effect in Estonia. Paper presented at the 10th Annual Meeting of the International Society for Intelligence Research, Madrid, Spain.
- Must, O., & Must, A. (December 13). Test-taking patterns have changed over time. Paper presented at the 13th Annual Meeting of the International Society for Intelligence Research, San Antonio, Texas.
- 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.
- Must, O., te Nijenhuis, J., Must, A., & van Vianen, A. E. M. (2009). Comparability of IQ scores over time. *Intelligence*, 37, 25–33.
- Neisser, U. (September–October). *Rising scores on intelligence tests*. American Scientist.
- Neisser, U. (1998). *The rising curve*. Washington: American Psychological Association (7).
- Nettelbeck, T., & Wilson, C. (2004). The Flynn effect: Smarter not faster. *Intelligence*, 32, 85–93.
- Rodgers, J. L., Cleveland, H. H., van den Oord, E., & Rowe, D. C. (2000). Resolving the debate over birth order, family size, and intelligence. *American Psychologist*, 55, 599–612.
- Rönnlund, M., & Nilsson, L. -G. (2008). The magnitude, generality, and determinants of Flynn effects on forms of declarative memory and visuospatial ability: Time-sequential analyses of data from a Swedish cohort study. *Intelligence*, 36, 192–209.
- Rönnlund, M., & Nilsson, L. -G. (2009). Flynn effects on sub-factors of episodic and semantic memory: Parallel gains over time and the same set of determining factors. *Neuropsychologia*, 47, 2174–2180.
- Rushton, J. P. (1985). Differential K theory: The sociobiology of individual and group differences. *Personality and Individual Differences*, 6, 441–452.
- Rushton, J. P. (1999). Secular gains in IQ not related to the g factor and inbreeding depression – Unlike Black–White differences: A reply to Flynn. *Personality and Individual Differences*, 26, 381–389.
- Rushton, J. P., & Ankney, C. D. (1996). Brain size and cognitive ability: Correlations with age, sex, social class, and race. *Psychonomic Bulletin & Review*, 3(1), 21–36.
- Rushton, J. P., & Jensen, A. R. (2010). The rise and fall of the Flynn effect as a reason to expect a narrowing of the Black–White IQ gap. *Intelligence*, 38, 213–219.
- Scarr, S., & Weinberg, R. A. (1978). The influence of “family background” in intellectual attainment. *American Sociological Review*, 43, 674–692.
- Shiu, W., Beaujean, A. A., Must, O., te Nijenhuis, J., & Must, A. (December 13). Item-level examination of the Flynn effect. Paper presented at the 13th Annual Meeting of the International Society for Intelligence Research, San Antonio, Texas.
- Smith, S. (1942). Language and nonverbal test performance of racial groups in Honolulu before and after a 14-year interval. *The Journal of General Psychology*, 26, 51–92.
- 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*, 33, 349–362.
- Sundet, J. M., Eriksen, W., Borren, I., & Tambs, K. (2010). The Flynn effect in sibships: Investigating the role of age differences between siblings. *Intelligence*, 38–1, 38–44.
- Tasbihsazan, R., Nettlebeck, T., & Kirby, N. (1997). Increasing mental development index in Australian children: A comparative study of two versions of the Bayley Mental Scale. *Australian Psychologist*, 32, 120–125.
- te Nijenhuis, J. T., Cho, S. H., Murphy, R., & Lee, K. H. (2012). The Flynn effect in Korea: Large gains. *Personality and Individual Differences*, 53, 147–151.
- te Nijenhuis, J. T., Murphy, R., & van Eeden, R. (2011). The Flynn effect in South Africa. *Intelligence*, 39(2011), 456–467.
- te Nijenhuis, J. T., van Vianen, A., & van der Flier, H. (2007). Score gains on g-loaded tests: No g. *Intelligence*, 35, 283–300.
- Teasdale, T. W., & Owen, D. R. (1989). Continuing secular increases in intelligence and a stable prevalence of high intelligence levels. *Intelligence*, 13, 255–262.
- Teasdale, T. W., & Owen, D. R. (2008). Secular declines in cognitive test scores: A reversal of the Flynn effect. *Intelligence*, 36, 121–126.
- Tuddenham, R. D. (1948). Soldier intelligence in World Wars I and II. *American Psychologist*, 3, 54–56.
- van Bloois, R. M., Geutjes, L. -L., te Nijenhuis, J., & de Pater, I. E. (December 19). g loadings and their true score correlations with heritability coefficients, giftedness, and mental retardation: Three psychometric meta-analyses. Paper presented at the Symposium on Group Differences, 10th Annual Meeting of the International Society for Intelligence Research, Madrid, Spain.
- Wai, J., & Putallaz, M. (2011). The Flynn effect puzzle: A 30-year examination from the right tail of the ability distribution provides some missing pieces. *Intelligence*, 39, 443–455.
- Wicherts, J. M., Dolan, C. V., Hessen, D., Oosterveld, P., Baal, G. C. M., van Boomsma, D. I., et al. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence*, 32, 509–537.
- Woodley, M. A. (2011). Heterosis doesn't cause the Flynn effect: A critical examination of Mingroni (2007). *Psychological Review*, 118(4), 689–693.
- Woodley, M. A. (2012). A life history model of the Lynn–Flynn effect. *Personality and Individual Differences*, 53(2), 152–156.