Intelligence 41 (2013) 753–764

Contents lists available at ScienceDirect

Intelligence

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.

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 "ΔIQ." A few typical national gains:

U.S. ΔIQ = 3 (14 points over 46 years, 1932–1978)
Estonia ΔIQ = 1.65 (12 points over 72 years, 1933/36 to 2006)
Japan ΔIQ = 7.7 (19 points over 25 years, 1940 to 1965)
Argentina ΔIQ = 6.91 (21.35 points over 34 years, 1964 to 1998).

E-mail address: bvv@cox.net.

0160-2896/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.intell.2013.04.010

[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 (ΔIQ = 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, Lluis-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	ΔIQ = 5.69	
Wechsler	ΔIQ = 5.2	
	Performance	ΔIQ = 7.8
	Verbal	ΔIQ = 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 (ΔIQ = 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 ΔIQ 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 ΔIQ of 1.65. But when the eight years from 1998 to 2006 were examined separately, the ΔIQ 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

ΔIQ = 3.63 Whites (same group took two different test batteries)

ΔIQ = 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