

“A Beautiful Theory, Killed by a Nasty, Ugly Little Fact”

A review of



Are We Getting Smarter? Rising IQ in the Twenty-First Century

by James R. Flynn

New York, NY: Cambridge University Press, 2012. 310 pp. ISBN

978-1-107-60917-4 (paperback). \$22.00, paperback

Reviewed by

[Alan S. Kaufman](#)

[Thomas Dillon](#)

[Jeffrey W. Kirsch](#)

James Flynn has a well-known effect named after him that posits an increase in a nation’s scores on IQ tests from one generation to the next. Within the United States, the *Flynn effect* is 3 points per decade and has been occurring at that steady rate since the 1930s for all ages. Many scientists agree that the Flynn effect is a valid and reliable phenomenon (e.g., Ceci & Kanaya, 2010; Kaufman, 2010a, 2010b), although few agree on what causes the effect.

In *Are We Getting Smarter? Rising IQ in the Twenty-First Century*, Flynn offers a theory to explain the effect that bears his name. Although the blurb inside the book proclaims “fascinating new material on a variety of topics,” much of the material is, in fact, old or rehashed. Flynn presented the same theory, and made many of the same arguments in

defense of his position, in his book *What Is Intelligence?* (Flynn, 2007, 2009); other “new” topics are from his articles published in 2009, 2010, and 2011.

His theory of intelligence to explain the Flynn effect reflects an integration of (a) individual differences on cognitive tasks, focusing on their cognitive complexity; (b) social adaptation to shifting social priorities; and (c) changes in the brain “localized neural clusters are developed differently as a result of specialized cognitive exercise” (p. 29). He focuses on the concept of *modernization* and viewing the world through “scientific spectacles” (p. 38). In a nutshell, he believes that we are scoring higher on IQ tests than previous generations did because “we live in a time that poses a wider range of cognitive problems than our ancestors encountered, and have . . . developed new cognitive skills and the kind of brains that can deal with these problems” (p. 27).

Remarkably, he argues his case with his own sense of forceful logic and idiosyncratic data but ignores a wealth of pertinent research in the fields of individual differences, neuropsychology, social psychology, cognitive neuroscience, and developmental psychology. For example, he compares scores of adults from different age cohorts tested on the Wechsler Adult Intelligence Scale—4th edition (WAIS-IV; Wechsler, 2008) and presents detailed tables of data about their relative performance. He refers to a *bright tax* (the penalty in “IQ points” for being intelligent) and comes to a variety of conclusions about aging—but he fails to integrate aging studies on the WAIS-IV (Lichtenberger & Kaufman, 2009; Salthouse & Saklofske, 2010) or its predecessors (Kaufman & Lichtenberger, 2006).

Yet Flynn’s most egregious error is in not acknowledging researchers who disagree with his conclusions about why the Flynn effect has occurred so steadfastly in the United States (and elsewhere) for nearly a century. A special issue of *Journal of Psychoeducational Assessment* (Kaufman & Weiss, 2010) devoted to the Flynn effect provided a scientific forum to debate psychometric and scientific issues such as Flynn’s interpretation of why IQs continue to rise, his contention that IQs should automatically be adjusted for the Flynn effect in capital punishment cases, his overreliance on data from specific subtests on Wechsler’s scales (most notably Similarities) and other homogeneous tests (Raven’s Progressive Matrices) to form the foundation of his theory, and so forth.

The special issue was a state-of-the-art compilation on the Flynn effect that included 11 articles by noted theorists, test developers, and cutting-edge researchers on the Flynn effect—including Flynn. Robert Sternberg stated, “We should not assume that because the name of a test stays the same, the actual constructs measured stay just the same as well” (Sternberg, 2010, p. 435). Lawrence Weiss, one of Flynn’s coauthors (Flynn & Weiss, 2007), concluded, “Changes in tests on revision complicate the measurement of rates of change” (Weiss, 2010, p. 491). Weiss also emphasized that Flynn’s grand notion of scientific spectacles “has invited numerous criticisms that must now be addressed—and that is as it should be in the spirit of scientific progress” (Weiss, 2010, p. 482).

Yet Flynn cites only one article from that 2010 special issue in *Are We Getting Smarter?*—his own (Flynn, 2010). He chose to continue to espouse his deeply held beliefs as

virtual fact while blatantly ignoring scientific and logical challenges to his theory. Flynn's theory is highly dependent on generational changes on two abilities among the many that are measured by IQ tests:

International Raven's data suggest that people have gained 50 points over the twentieth century. It has one rival. . . . One [Wechsler] subtest shows gains near the magnitude of Raven's gains. It is the Similarities subtest, which tests your ability to classify things that have something in common (e.g., dogs and rabbits are both mammals). (p. 7)

Flynn's theory focuses on the increased ability of people of today to respond to logic problems with more abstractness than did our largely concrete forefathers—as proved by the huge improvement over time on Raven's and Similarities tests. Never mind that Kaufman and others in the special issue challenged the legitimacy of comparing scores on these tasks across generations.

For example, Kaufman (2010a) noted that the directions for administering and scoring Similarities changed drastically from the 1949 Wechsler Intelligence Scale for Children (WISC) to the 1974 WISC-R (and from the 1955 WAIS to the 1981 WAIS-R). Among the many changes:

The WISC never gave children a clue that abstract answers were worth more than concrete answers. The scoring system awarded 2 points for abstract categories and 1 point for functions or descriptive qualities, but children were never given appropriate feedback to shape their responses from concrete to abstract. That changed on the WISC-R. On Item 5 [e.g., How are a PLUM and PEACH alike?], examiners were instructed to tell children who gave a 1-point response (e.g., "You eat them both"), "That's right. You do eat them both. Also, they both are fruits." (p. 385)

It is feasible that the huge gains on Similarities over time, which Flynn uses to support his contention that present-day Americans wear "scientific spectacles" and are able to respond much more abstractly than our ancestors did, are at least partly due to the change in test directions that encourage children and adults to respond more abstractly. Others in the special issue agreed with Kaufman's notion that comparing Similarities test scores over time is of questionable validity (Ceci & Kanaya, 2010; McGrew, 2010; Sternberg, 2010) or at least agreed that the topic is worthy of scientific study (Weiss, 2010).

Flynn (2010) disagreed, and he has ignored this relevant debate in *Are We Getting Smarter?* His simple, unchallenged conclusion, in total support of his theory, is "Then there is similarities, whose gains are a result of more and more people donning scientific spectacles" (p. 38).

Comparing scores on Raven's matrices over a century is also questionable. Indeed, Flynn, in the book under review, provides evidence of societal changes that would argue

against the validity of comparing Raven's scores from one generation to another. He shows how the fluid, nonschool-like Raven's items were becoming part of the curriculum in American schools because of the relationship between mathematical thinking and Raven's problem-solving ability: "Therefore, it appears sensible to teach young children Raven's-type problems in the hope that they will become better mathematics problem solvers. US schools have been doing that since 1991 (Blair *et al.*, 2005, pp. 100–101)" (p. 17).

Nonetheless, Raven's data are sliced and diced to argue against "nutrition" as a competing argument for explaining the Flynn effect (pp. 40–52). In the course of his data juggling and questionable assumptions, he never actually cites a study that specifically examined the relationship of nutrition to IQ gain, but he does make outlandish statements based on numerous comparisons of different age cohorts on Raven's—"either some dramatic improvement in lower-class diet set in about 1973; or some dramatic deterioration set in about 2001" (p. 50).

The section on nutrition is illustrative of his arbitrary use, and sometimes abuse, of data. His extrapolations of the Flynn effect illustrate this tendency: "In 1917, Americans had a mean IQ of 72 (against today's norms) and a good estimate for 1900 would be 67" (p. 32). Binet and Simon didn't even develop the first IQ test until after 1900 (Binet & Simon, 1904)! More offensive, however, are his headings for various countries that proclaim their mean IQ to be in ranges that clinicians refer to as *borderline* or what used to be called *dull normal* (e.g., Kenya, IQ = 72; Saudi Arabia, IQ = 84; Dominica, IQ = 82; Sudan, IQ = 71).

In general, the book is disorganized, rambles from topic to topic, offers no overview or guiding principle, and treats data in an arbitrary and haphazard manner to test this or that hypothesis. Sometimes Flynn's comments read more like astrology than psychology. He says the following about massive gains on Raven's: "Taking hypothetical situations seriously may have rendered moral debate more reflective" (p. 23). Regarding adult gains on Wechsler's Vocabulary subtest, he says the following: "[S]erious writers today have a larger target audience capable of reading their works, although the visual culture of our time may limit the number of those willing to do so" (pp. 21–22).

The text is filled with acronyms, jargon, dates, and citations. The many syllogisms are seldom closed in the body of the text. The style is to offer a rash of data and citations and follow them with an assertion that flows more from the author's convictions than from logic or the findings of sound empirical research.

Most important, the book fails to use the Flynn effect to shed light on human intelligence and the deeper meaning of IQ testing. It seems clear that the Flynn effect exists and in some cases causes truly dramatic changes in IQ results. However, at the end of a careful reading, we see it as a more complex issue rather than a simple one. Ultimately, *Are We Getting Smarter?* fails to deliver on the promise of its title and its introductory chapter.

The book is targeted for psychologists and the bright lay public as well. It has been discussed and accepted fairly uncritically in a variety of mainstream publications such as *Science* (see review by Schleicher, 2013), the *APA Monitor on Psychology* (see interview

with Flynn by Winerman, 2013), the *Wall Street Journal* (Flynn, 2012), and the *New York Times* (Kristof, 2012). To offer multiple perspectives on the book, this entire review has been written jointly by a psychologist who testified on the same side as Flynn in a Texas capital punishment law case in 2010 (Alan S. Kaufman) and by two professionals who are not in the field: a physical chemist who worked as an engineering executive (Thomas Dillon) and an aerospace engineer who is a museum director and film producer (Jeffrey W. Kirsch).

Alan S. Kaufman is astonished that Flynn based his myriad computations of the WAIS Flynn effect using 2006 as the WAIS–IV standardization date when, in fact, WAIS–IV data were collected between March 2007 and April 2008 (Diane L. Coalson, personal communication, October 8, 2008). Thomas Dillon and Jeffrey W. Kirsch observe that although the text is somewhat accessible to a nonpsychologist, it is frequently an exhausting, even painful, read. The author almost never alerts the reader at the beginning of a chapter, section, or paragraph what he will prove. The chapters on youth and age as well as race and gender are easier to follow and much more compelling for a nonpsychologist.



In addition, Dillon and Kirsch note that any understanding of the relationship between “smarter” and IQ test results has been complicated much more than clarified. The book could not be recommended for a person not in the field. As outsiders looking in on the field of statistical analysis of intelligence testing, they are amazed that from the perspective of Flynn (and presumably his adherents), there has been no serious challenging of the Flynn effect. Unless the reader is very familiar with the *g* theory/model and its significance, everything else seems to be too strongly based on Flynn’s individual assessment. Is he fairly representing fieldwide accepted interpretations? Or does he have a very individualistic understanding of the test results and their implications?


Ultimately, all three of us agree that Flynn’s interpretation is individualistic and steadfast in the face of alternative perspectives and data that might be inconsistent with his hypotheses. We are reminded of Francis Galton’s recounting of a dinner conversation between Herbert Spencer and Thomas H. Huxley in which Spencer asserted that he once wrote a tragedy, and Huxley insisted that he already knew of the plot: “A beautiful theory, killed by a nasty, ugly little fact” (Galton 1908, p. 258; see also Huxley, 1870/1896).

References



- Binet, A., & Simon, T. (1904). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux [New methods for the diagnosis of abnormal intellectual level]. *L'Année Psychologique*, *11*, 191–244. doi:10.3406/psy.1904.3675
- 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. doi:10.1016/j.intell.2004.07.008  

Ceci, S. J., & Kanaya, T. (2010). “Apples and oranges are both round”: Furthering the discussion on the Flynn effect. *Journal of Psychoeducational Assessment*, 28, 441–447. doi:10.1177/0734282910373339  

Flynn, J. R. (2007). *What is intelligence?* New York, NY: Cambridge University Press. 

Flynn, J. R. (2009). *What is intelligence?* (2nd ed.). New York, NY: Cambridge University Press.



Flynn, J. R. (2010). Problems with IQ gains: The huge vocabulary gap. *Journal of Psychoeducational Assessment*, 28, 412–433. doi:10.1177/0734282910373342  

Flynn, J. R. (2012, September 21). Are we really getting smarter? *The Wall Street Journal*. Retrieved from http://online.wsj.com/article/SB10000872396390444032404578006612858486012.html?mod=WSJ_hpp_MIDDLENexttoWhatsNewsForth

Flynn, J. R., & Weiss, L. (2007). American IQ gains from 1932 to 2002: The WISC subtests and educational progress. *International Journal of Testing*, 7, 209–224. doi:10.1080/15305050701193587  



Galton, F. (1908). *Memories of my life*. London, England: Methuen.







Huxley, T. H. (1896). Biogenesis and abiogenesis. In T. H. Huxley, *Discourses: Biological and geological essays* (pp. 229–271). New York, NY: D. Appleton. (Original work published 1870)

Kaufman, A. S. (2010a). “In what way are apples and oranges alike?”: A critique of Flynn’s interpretation of the Flynn effect. *Journal of Psychoeducational Assessment*, 28, 382–398. doi:10.1177/0734282910373346  

Kaufman, A. S. (2010b). Looking through Flynn’s rose-colored scientific spectacles. *Journal of Psychoeducational Assessment*, 28, 494–505. doi:10.1177/0734282910373573  

Kaufman, A. S., & Lichtenberger, E. O. (2006). *Assessing adolescent and adult intelligence* (3rd ed.). New York, NY: Wiley. 

Kaufman, A. S., & Weiss, L. G. (Eds.). (2010). Guest editors’ introduction to the special issue of *Journal of Psychoeducational Assessment* on the Flynn Effect. *Journal of Psychoeducational Assessment*, 28, 379–381. doi:10.1177/0734282910373344  

- Kristof, N. D. (2012, December 12). It's a smart, smart world [Op-ed column]. *The New York Times*. Retrieved from <http://www.nytimes.com/2012/12/13/opinion/kristof-its-a-smart-smart-smart-world.html?ref=opinion&pagewanted=print>
- Lichtenberger, E. O., & Kaufman, A. S. (2009). *Essentials of WAIS-IV assessment*. Hoboken, NJ: Wiley.
- McGrew, K. S. (2010). The Flynn effect and its critics: Rusty linchpins and “lookin’ for g and Gf in some of the wrong places.” *Journal of Psychoeducational Assessment*, 28, 448–468. doi:10.1177/0734282910373347  
- Salthouse, T. A., & Saklofske, D. H. (2010). Do the WAIS-IV tests measure the same aspects of cognitive functioning in adults under and over age 65? In L. G. Weiss, D. H. Saklofske, D. L. Coalson, & S. E. Raiford (Eds.), *WAIS-IV clinical use and interpretation* (pp. 217–235). San Diego, CA: Academic Press. doi:10.1016/B978-0-12-375035-8.10008-4
- Schleicher, A. (2013, January 25). Making sense of rising IQ scores. *Science*, 339, 394–395. doi:10.1126/science.1232678
- Sternberg, R. J. (2010). The Flynn effect: So what? *Journal of Psychoeducational Assessment*, 28, 434–440. doi:10.1177/0734282910373349  
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). San Antonio, TX: Pearson.
- Weiss, L. G. (2010). Considerations on the Flynn effect. *Journal of Psychoeducational Assessment*, 28, 482–493. doi:10.1177/0734282910373572  
- Winerman, L. (2013, March). Questionnaire: Smarter than ever? *APA Monitor on Psychology*, 44(3), 30–33.
-