DOI: 10.1111/cdev.13675

EMPIRICAL ARTICLE

CHILD DEVELOPMENT

Genetically informed, multilevel analysis of the Flynn Effect across four decades and three WISC versions

Evan J. Giangrande[1] | **Christopher R. Beam**[2] | **Deborah Finkel**[3,4] | **Deborah W. Davis**[5] | **Eric Turkheimer**[1]

[1]Department of Psychology, University of Virginia, Charlottesville, Virginia, USA
[2]Department of Psychology, University of Southern California, Los Angeles, California, USA
[3]Department of Psychology, Indiana University Southeast, New Albany, Indiana, USA
[4]Institute for Gerontology, Jönköping University, Jönköping, Sweden
[5]Department of Pediatrics, University of Louisville School of Medicine, Louisville, Kentucky, USA

Correspondence
Evan J. Giangrande, Department of Psychology, University of Virginia, PO Box 400400, Charlottesville, VA 22904-4400, USA.
Email: ejg2br@virginia.edu

Funding information
This work was supported by the National Institute on Aging (R01AG063949-01 and R03AG048850-01).

Abstract

This study investigated the systematic rise in cognitive ability scores over generations, known as the *Flynn Effect*, across middle childhood and early adolescence (7–15 years; 291 monozygotic pairs, 298 dizygotic pairs; 89% White). Leveraging the unique structure of the Louisville Twin Study (longitudinal data collected continuously from 1957 to 1999 using the Wechsler Intelligence Scale for Children [WISC], WISC–R, and WISC–III ed.), multilevel analyses revealed between-subjects Flynn Effects—as both decrease in mean scores upon test re-standardization and increase in mean scores across cohorts—as well as within-child Flynn Effects on cognitive growth across age. Overall gains equaled approximately three IQ points per decade. Novel genetically informed analyses suggested that individual sensitivity to the Flynn Effect was moderated by an interplay of genetic and environmental factors.

In 1984, James Flynn found that American standardization samples, including children, scored systematically higher on older versions of IQ tests than they did on newer versions, reflecting a 13.8-point rise in mean IQ scores between 1932 and 1978 (approximately three points per decade; Flynn, 1984). The secular rise in standardized intelligence scores, now known as the *Flynn Effect* (FE), has since been replicated widely in children and adults (Pietschnig & Voracek, 2015). The FE is typically documented by giving the same sample two versions of a cognitive ability test and noting lower mean scores on newer versions than on older versions, as in Flynn's original study (we refer to these as *test version effects*). Other evidence comes from studies of European military conscripts, in which cohorts from more recent generations scored systematically higher than previous cohorts on the same test version (*cohort effects*; e.g., Sundet et al., 2004; Teasdale & Owen, 2008).

Meta-analytic estimates have been roughly consistent with Flynn's original observation of an increase in three IQ points per decade, with measures of *fluid intelligence* (problem solving and abstract reasoning that do not rely on previous knowledge; measures include performance IQ [PIQ]) frequently showing greater gains over time than measures of *crystallized intelligence* (application of knowledge previously acquired through experience and education; measures include verbal IQ [VIQ] and vocabulary; Pietschnig & Voracek, 2015). Studies from

Abbreviations: DOB, date of birth; DZ, dizygotic; FE, Flynn Effect; FSIQ, full-scale IQ; LTS, Louisville Twin Study; MZ, monozygotic; PIQ, performance IQ; SES, socioeconomic status; VIQ, verbal IQ; WISC, Wechsler Intelligence Scale for Children; WISC–III, Wechsler Intelligence Scale for Children, 3rd ed.; WISC–R, Wechsler Intelligence Scale for Children–Revised.

developing regions tend to yield even larger effects (e.g., children in rural Kenya; Daley et al., 2003), whereas IQ gains in developed regions appear to have decelerated over the past century (Pietschnig & Voracek, 2015), and evidence suggests that scores have even started to decline slightly (Sundet et al., 2004; Teasdale & Owen, 2005, 2008). These results suggest that the FE is positively linked to societal modernization and development; scores appear to rise as regions develop until possibly reaching an asymptote.

The FE has had a profound impact on developmental researchers' conceptualizations of cognitive ability. It provides evidence that broad changes to environments in which children are raised can influence cognitive development, which is driven by a dynamic interplay of genetic *and* environmental factors (Dickens & Flynn, 2001; Flynn, 2007; Pietschnig & Voracek, 2015). Findings regarding the relative strength of the FE at different stages of development have been inconsistent, with one meta-analysis observing larger effects in adults than in children (Pietschnig & Voracek, 2015) and another finding no differences across ages (Trahan et al., 2014).

Practical implications

In addition to informing theoretical perspectives on cognitive development, the FE has altered the day-to-day lives of individual children. Howard (2001) provided correlational evidence of real-world changes that may reflect rising cognitive ability. More disconcertingly, the FE influences outcomes that depend on falling above or below a set IQ score cutoff, especially for children. In the United States, intellectual disability is commonly identified in part as having an IQ score below 70 (e.g., American Psychiatric Association, 2013). Work by Kanaya, Ceci, and Scullin has demonstrated that rates of intellectual disability diagnosis drop steadily in the years prior to the introduction of a new IQ test version; due to the

and achievement scores, significantly reducing the probability that a child previously diagnosed with learning disability using an old test version will be re-diagnosed. This creates an opposite problem to that seen in intellectual disability: children who were once enrolled in special education programs become ineligible. Abruptly losing special education resources can be extremely difficult for children, forcing them to grapple with academic and social challenges without the supports on which they formerly relied.

Research implications

The FE poses a serious problem in cross-sectional studies of cognitive ability that include multiple generational cohorts or multiple test versions. In one striking example, the FE accounted for about 85% of supposedly age-related IQ disparities between 20- and 70-year olds (Dickinson & Hiscock, 2010). Even when comparing groups at a single age, failing to adjust for between-group differences in cohort or test version could make groups appear more dissimilar than they really are.

The FE may also confound longitudinal studies in which individuals take multiple IQ test versions, such as studies of cognitive growth in children. Consider a longitudinal analysis in which two different versions of the Wechsler Intelligence Scale for Children (WISC; commonly used to measure cognitive ability in children) were administered at ages 8 and 16. If intra-individual changes in cognitive ability are observed over that span, how can one be sure that they reflect actual cognitive changes and not test version artifacts introduced by the FE? If changes are *not* observed, could test re-standardization have flattened effects that otherwise would have been observed? The FE could be particularly problematic in analyses of unstandardized raw scores, which change substantially as children grow intellectually.