Intelligence 77 (2019) 101385

Contents lists available at ScienceDirect

Intelligence

journal homepage: www.elsevier.com/locate/intell

The Flynn effect for fluid IQ may not generalize to all ages or ability levels: A population-based study of 10,000 US adolescents

Jonathan M. Platt[a,*], Katherine M. Keyes[a,b], Katie A. McLaughlin[c], Alan S. Kaufman[d]

[a] Department of Epidemiology, Mailman School of Public Health, Columbia University, 722 West 168th Street, New York, NY 10032, United States of America
[b] Center for Research on Society and Health, Universidad Mayor, Santiago, Chile
[c] Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, United States of America
[d] Yale University, Child Study Center, School of Medicine, 230 S. Frontage Road, New Haven, CT 06519, United States of America

ARTICLE INFO

Keywords:
Intelligence
Flynn effect
Adolescence
Intellectual disabilities

ABSTRACT

Generational changes in IQ (the Flynn Effect) have been extensively researched and debated. Within the US, gains of 3 points per decade have been accepted as consistent across age and ability level, suggesting that tests with outdated norms yield spuriously high IQs. However, findings are generally based on small samples, have not been validated across ability levels, and conflict with reverse effects recently identified in Scandinavia and other countries. Using a well-validated measure of fluid intelligence, we investigated the Flynn Effect by comparing scores normed in 1989 and 2003, among a representative sample of American adolescents ages 13–18 (n = 10,073). Additionally, we examined Flynn Effect variation by age, sex, ability level, parental age, and SES. Adjusted mean IQ differences per decade were calculated using generalized linear models. Overall the Flynn Effect was not significant; however, effects varied substantially by age and ability level. IQs increased 2.3 points at age 13 (95% CI = 2.0, 2.7), but decreased 1.6 points at age 18 (95% CI = −2.1, −1.2). IQs decreased 4.9 points for those with IQ ≤ 70 (95% CI = −4.9, −4.8), but increased 3.5 points among those with IQ ≥ 130 (95% CI = 3.4, 3.6). The Flynn Effect was not meaningfully related to other background variables. Using the largest sample of US adolescent IQs to date, we demonstrate significant heterogeneity in fluid IQ changes over time. Reverse Flynn Effects at age 18 are consistent with previous data, and those with lower ability levels are exhibiting worsening IQ over time. Findings by age and ability level challenge generalizing IQ trends throughout the general population.

1. Introduction

Societal changes in intelligence from generation to generation have been studied for nearly a century (Rundquist, 1936), both in the US (Doppelt & Kaufman, 1977; Flynn, 1984) and cross-nationally (Flynn, 1987; Pietschnig & Voracek, 2015). While there is wide variation in the magnitude of generational change in IQ across countries (Pietschnig & Voracek, 2015), in the US children and adults score higher on IQ tests than previous generations at the rate of approximately 3 IQ points per decade. It has been argued that this rate of change has been true for about a century (Pietschnig & Voracek, 2015; Trahan, Stuebing, Fletcher, & Hiscock, 2014). Such generational gains have become known as the *Flynn Effect*, after the scientist who popularized the concept (Flynn, 1984, 2012). As a direct consequence of the rising IQs over generations, standards or reference groups for IQ tests (norms) become out of date at the rate of 3 IQ points per decade in the US. These "soft" norms yield IQs that are spuriously inflated by the Flynn Effect; people who earn IQs of 85 (for example) on a test that was standardized 10 years ago have been given an artificial boost of 3 IQ points such that their "true" IQ is 82 when adjusted for the outdated norms. Indeed, guidelines for the *American Association on Intellectual and Developmental Disabilities* (AAIDD) (McGrew, 2015; Schalock, 2012) advise subtracting 3 points per decade from the obtained IQ whenever a test with outdated norms has been administered to someone suspected of an intellectual disability.

However, recent research suggests the Flynn Effect may be changing. Since the end of the 20th Century, studies have found a *reverse* Flynn Effect among birth cohorts of very large samples of young adult males in Norway (Bratsberg & Rogeberg, 2018; Sundet, Barlaug, & Torjussen, 2004), Denmark (Teasdale & Owen, 2005, 2008), and Finland (Dutton & Lynn, 2013). Reverse Flynn Effects have also been reported in several other countries (Dutton, van der Linden, & Lynn,

* Corresponding author.
E-mail address: jmp2198@cumc.columbia.edu (J.M. Platt).

https://doi.org/10.1016/j.intell.2019.101385
Received 11 February 2019; Received in revised form 21 August 2019; Accepted 23 August 2019

2016), although data for some countries are based on non-traditional measures of IQ (Piagetian tests in Great Britain), group-administered aptitude tests (Netherlands), measures of spatial perception (German-speaking countries), or on small sample sizes (two groups of 79 adults in France). Despite these caveats, previous data suggest that IQs may be decreasing in recent birth cohorts, especially in Europe. It is unknown whether a similar reversal of the Flynn Effect occurred in the US during this time.

Additional questions regarding the Flynn Effect include whether it is consistent across age and the entire range of intellectual abilities. The validity of adjustments in IQ among those at the lower ends of IQ has major implications for placement and eligibility for educational services. The largest study of this question included nearly 9000 children and adolescents (ages 6–17) across nine US school systems from 1989 to 1995 studied before and after schools transitioned from one intelligence test for children (i.e., the Wechsler) to its revision and found evidence of the Flynn Effect at the low IQ range (Kanaya, Scullin, & Ceci, 2003). Students with IQs below 80 were more likely to be classified as intellectually disabled based on the revised test (with newer, steeper norms) than when tested on the older version. Other studies have addressed the question, but with methodologies that varied widely in quality (e.g., small sample sizes of children in the low range of IQ) (Zhou, Gregoire, & Zhu, 2010) and sometimes with atypical samples such as children with learning disabilities (Sanborn, Truscott, Phelps, & McDougal, 2003).

Finally, studies of the Flynn Effect are typically based on comparisons of individuals tested with two different instruments—often an IQ test and its modified revision (Pietschnig & Voracek, 2015; Trahan et al., 2014). These studies introduce unwanted error due to practice effects (test experience) and different sets of items administered on each test. Overall, these limitations have contributed to an inconsistent and inconclusive understanding of the Flynn Effect in recent years (McGrew, 2015; Spitz, 1989; Trahan et al., 2014; Weiss, 2010).

In light of these limitations, the present study investigated the consistency and patterns of the Flynn Effect in recent decades in the US, using a nationally representative sample of > 10,000 adolescents, ages 13–18, who completed the nonverbal portion of the Kaufman Brief Intelligence Test (K-BIT) (Kaufman & Kaufman, 1990). K-BIT scores were normed in 2001-04 then compared with scores from a set of norms developed in 1988–1989. The use of two sets of norms, rather than two separate tests, avoids potential methods-based bias. The goals of this study were: (a) to determine whether the Flynn Effect (or a reverse effect) was observed for fluid IQ in the US during the time frame that yielded reverse Flynn Effects in several European countries; (b) to determine whether the Flynn Effect was consistent across intellectual ability levels, especially in the IQ range associated with intellectual disability (IQ ≤ 70 ± 5); and (c) to examine the relationship of the Flynn Effect to a variety of socio-demographic variables to examine potential heterogeneity in the effect.

2. Materials and methods

2.1. Sample

Data were from the National Comorbidity Survey Adolescent Supplement (NCS-A), a US population-representative study of the prevalence and correlates of psychiatric disorders in adolescence (age range: 13–18 years). In total, 10,148 adolescents were recruited from both schools and households and completed the survey between 2001 and 2004 (Kessler et al., 2009). Of these, 10,073 (99.3%) completed a supplemental survey that included an individually administered measure of fluid intelligence, described below. Poststratification weighting adjusted for minor differences in sample and population distributions of 2000 census socio-demographic and school frequencies (Kessler, Avenevoli, Costello, et al., 2009). Parents/guardians gave written informed consent and adolescent participants gave written informed assent after receiving a complete description of the study, in accordance to the procedures approved by Human Subjects Committees of Harvard Medical School and the University of Michigan. The Institutional Review Board of Columbia University approved the present analysis (IRB-AAAN1104). Study participants were compensated $50 for participation. Additional study details have been published (Kessler et al., 2009). The analytic sample included those with valid K-BIT data and non-missing survey weights (n = 10,073).

2.2. Fluid intelligence

Intelligence was measured using the 48-item nonverbal portion of the Kaufman Brief Intelligence Test (K-BIT), a standardized measure of fluid intelligence (Kaufman & Kaufman, 1990; Kaufman & Wang, 1992). The K-BIT uses abstract matrices similar to those developed by Raven (1936), which have become widely accepted as prototypical measures of fluid reasoning (Kaufman, 2009)—a cognitive ability that relates closely to psychometric *g* in factor analyses (Floyd, Reynolds, Farmer, & Kranzler, 2013; Reynolds, Floyd, & Niileksela, 2013).

The K-BIT was administered by non-clinical interviewers who received appropriate training and practice, in accordance to the original administration procedures (Bain & Jaspers, 2010; Kaufman & Kaufman, 1990). The K-BIT nonverbal sections have strong internal consistency (range: 0.87–0.92) and adequate test-retest reliability (range: 0.76–0.89 (Kaufman & Kaufman, 1990; Salthouse, 2010), including in the present sample (alpha = 0.90). The instrument has demonstrated invariance by sex and ethnicity and has established good construct validity with theory-based and other established measures of intelligence throughout adolescence (Bain & Jaspers, 2010; Canivez, Neitzel, & Martin, 2005; Homack & Reynolds, 2007; Kaufman, Johnson, & Liu, 2008; Kaufman & Kaufman, 1990; Kaufman & Wang, 1992; Wang & Kaufman, 1993). Additional details regarding test administration and scoring have been previously published (Keyes, Platt, Kaufman, & McLaughlin, 2016). Using the 2003 K-BIT scores in the NCS-A sample, we have previously examined the relationship of IQ with: psychiatric disorders (Keyes et al., 2016), childhood adversity (Platt et al., 2018), and as a component of intellectual disability (Platt, Keyes, McLaughlin, & Kaufman, 2018); this is the only study using 1989 scores as well and the only one that has examined the Flynn Effect.

From the K-BIT raw scores, updated norms were created specifically for the NCS-A by the test developer. Scores were normed within six-month age groups (hereafter referred to as 2003 K-BIT scores). Original K-BIT norms were developed during a national standardization procedure. Between 1988 and 1989 a US population-representative sample of 2022 subjects ages 4–90 years were recruited from 60 sites nationwide and adolescent K-BIT scores were age-normed by six-month intervals (hereafter referred to as 1989 K-BIT scores) (Kaufman & Kaufman, 1990). The 1989 norms were stratified, based on Census estimates in 1985, 1987, and 1990, on age, sex, parental education, race or ethnic group, and geographic region. The adolescent portion of the norming sample included individuals ages 13–14 (n = 181), 15–16 (n = 148), and 17–19 (n = 180) (Kaufman & Kaufman, 1990, Tables 4.1–4.6). The 2003 norms matched Census data for 2000 frequencies for a wide array of sociodemographic, economic, and other background variables such as age, sex, parent education, race, ethnicity, urbanicity, parent education, and parent income level.

Based on the precise midpoints of the dates of administration, 13.67 years elapsed between data collection for the two normative samples. The Flynn Effect is defined as the rate of gain (or loss) of IQ points over time on the same, or similar, instruments. The two standardizations of the K-BIT nonverbal test with representative samples, obtained more than a decade apart, meet the precise criteria required to assess the Flynn Effect within the US.

2.3. Socio-demographic and environmental variables

Flynn Effects were stratified by the following socio-demographic and environment characteristics: age (range 13–18), sex, parent education (less than a HS degree, HS graduate, some college, college degree or more), parent income level (< 1.5, 1.5–3, 3.1–6, > 6 times the poverty level), birth order (range 1–5), number of siblings (0–5 or more), and maternal/paternal age when the respondent was born (range: < 20–40 or older. Ability level was defined as a categorical variable of sequential IQ levels (e.g., ≤70, 71–75, 76–79, 80–89, 90–99, 100–109, 110–119, 120–124, 125–129, 130 +). Groups were defined using 1989 norms, as they yielded larger sample sizes among groups with very low and very high levels; however, Flynn Effect estimates based on 2003 norms were not meaningfully different (Supplementary Table 2).

3. Analysis

To calculate the Flynn Effect, we examined the difference in each participant's IQ as calculated from the norms developed in 1989 and in 2003, subtracting the IQ as estimated using norms developed in 2003 from the IQ based on norms developed in 1989, and reported as IQ change per decade (i.e., (IQ difference/13.67)*10). We calculated 95% confidence intervals (CIs) based on the sample distribution of mean IQ differences. Effect sizes were also calculated by dividing the adjusted mean difference by the overall sample standard deviation (SD = 15), in order to facilitate between-group comparison of the magnitude of Flynn Effects.

Based on the Flynn Effect, the expectation is that any person's IQ on the older 1989 norms would be a few points higher than his or her IQ calculated on the newer 2003 norms. If Flynn Effect theory is correct, then the older 1989 norms have become outdated by about 14 years relative to the newer 2003 norms; hence the old norms are "soft" and spuriously inflate IQs of adolescents tested almost a decade and a half later. For each individual in our sample, a positive IQ difference (IQ based on 1989 norm higher than IQ based on 2003 norm) denotes a positive Flynn Effect, consistent with theory and previous meta-analyses. By contrast, a negative IQ difference denotes a *reverse* Flynn Effect, consistent with the findings of IQ loss among Scandinavian military conscripts between the 1990s and early 2000s.

Flynn Effects were estimated using generalized linear models incorporating sample weight, adjusted for the socio-demographic and environmental variables described above, as well as: race/ethnicity (Hispanic Non-Hispanic Black, Non-Hispanic White, other), parent nationality (born in the US, not born in the US), any psychiatric disorder, and any substance use disorder. Estimates were adjusted in order to minimize any residual differences between the 1989 and 2003 norms samples that may cause confounding.

We examined differences in Flynn Effects within all socio-demographic groups described above. We did not test for additive statistical interaction these associations, because statistical significance does not always connote meaningful differences (especially with very large sample sizes). Rather, we focused where within-group Flynn Effects are appreciably different with reasonable confidence intervals. In addition, where we found within-group differences, we investigated whether findings differed further by sex, age group (dichotomized as younger ages = 13–15 and older ages = 16–18), and 10-category ability level.

4. Results

The adjusted mean K-BIT scores and mean change per decade (i.e., Flynn Effects) are presented in Table 1. Overall, there was no evidence of the Flynn Effect in the total sample (M = 0.09; 95% CI = 0.05, 0.12). However, there was meaningful variation in the Flynn Effect by level of age and intellectual ability. The Flynn Effect decreased monotonically with increasing age. Respondents who were age 13 had a mean IQ increase of 2.3 points per decade (95% CI = 2.0, 2.7), whereas respondents who were age 18 had a *decrease* of 1.6 IQ points per decade (95% CI = -2.1, −1.2). Across intellectual ability levels, IQ decreased 4.9 points per decade for those with IQ ≤ 70 (95% CI = -4.9, −4.8), and increased approximately linearly to 3.5 points for those with IQ ≥ 130 (95% CI = 3.4, 3.6). The Flynn Effect did not differ by sex or as a function of any other socio-demographic or environmental variables.

Next, we conducted several stratified analyses to explore sex differences in Flynn Effects by age and ability level, and age differences in Flynn Effects by ability level. Fig. 1 summarizes the adjusted mean differences in IQ assessed between 1989 and 2003 norms by ability level, stratified by sex. The negative values at low ability levels suggest reverse Flynn Effects; that is, those at low ability levels test *lower* using 1989 norms than they would have using 2003 norms. In contrast, those at high ability levels test *higher* using 1989 norms than they would have using 2003 norms. These differences were consistent across sex. Supplementary Table 3 presents these estimates numerically. For those with IQ ≤ 70, there was a reverse Flynn effect for both females (−5.0; 95% CI = −5.1, −4.9) and males (−4.7; 95% CI = 4.8, 4.6). For those with IQ ≥ 130, there is a positive Flynn effect for both females (3.8; 95% CI = 3.6, 4.0) and males (3.0; 95% CI = 3.2, 3.5).

Fig. 2 summarizes the adjusted mean differences in IQ assessed between 1989 and 2003 norms by age, stratified by sex. The overall age effects we describe above were consistent between males and females. Fig. 3 and Supplementary Table 4 summarize the adjusted mean differences in IQ assessed between 1989 and 2003 norms by ability level stratified by age group. Both age groups displayed reverse Flynn Effects for adolescents with IQ ≤ 70: (ages 13–15 FE = -4.4; 95% CI = −4.5, −4.3); ages 16–18 FE = −5.4; 95% CI = −5.5, −5.4). However, at higher IQs, the positive Flynn Effect was much more notable for ages 13–15 (4.4; 95% CI = 4.2, 4.5) than for 16–18 (1.6; 95% CI = 1.5, 1.8). In fact, most IQ changes for the younger group were positive, with values of 2 points or more per decade characterizing those with IQs 120 or greater. By contrast, most IQ changes for ages 16–18 were negative, with reverse Flynn Effects of 2 points per decade or greater for those with IQs < 80. This finding indicates an interaction between age and ability level whereby the reverse Flynn Effect associated with low IQs was more pronounced for older adolescents and the positive Flynn Effect for those with high IQs was more robust for younger adolescents.

To explore this interaction further, we analyzed data for more homogeneous sub-samples, computing Pearson product-moment correlations between each adolescent's K-BIT IQ and their individual Flynn Effect. We conducted these analyses separately by age group because of the different Flynn Effects found by age using IQs derived from the 1989 and 2003 norms. Coefficients were substantial and generally consistent between norms, with values ranging from 0.51 to 0.83 (mean $r = 0.61$) using 2003 norms and 0.73–0.85 (mean $r = 0.77$) using 1989 norms, indicating a positive correlation between IQ and the magnitude of the Flynn Effect.

5. Discussion

The present study utilized data from a large US-representative sample of adolescents to describe changes in IQ between 1989 and 2003. There were three central findings: 1) Overall, there was no evidence of a Flynn Effect during the study period; 2) however, overall IQ trends masked substantial heterogeneity in the presence and direction of the Flynn Effect by both ability level and age; and 3) there was no variation in the Flynn effect as a function of other sociodemographic characteristics.

The overall lack of a Flynn Effect in our sample is concordant with trends in the K-BIT, KBIT-2, the Kaufman Assessment Battery for Children (K-ABC and KABC-II), and other individually administered screening tests reported in a previous meta-analysis (Trahan et al., 2014). It also conforms with the conclusion that gains have decreased

Table 1
Adjusted mean K-BIT scores and mean change, comparing norms set in 1989 with norms set in 2003 (Flynn Effects), in a nationally-representative sample of US adolescents—IQ groupings based on 1989 norms.

	N	1989 K-BIT norms			2003 K-BIT norms			Mean K-BIT difference		Flynn effect[*,†]			Effect size[‡]
		Mean	95% CL		Mean	95% CL		Unadjusted	Adjusted[*]		95% CL		
Total	10,073	100.4	99.6	101.2	100.3	99.5	101.1	0.13	0.12	0.09	0.05	0.12	0.01
Age group													
13	1547	104.1	102.5	105.7	100	98.6	101.4	4.02	3.19	2.33	1.97	2.69	0.21
14	2234	102.3	100.8	103.7	100.6	99.2	101.9	1.66	1.2	0.88	0.67	1.09	0.08
15	1886	99.2	97.3	101.1	100.3	98.6	102	−1.10	−1.22	−0.89	−0.97	−0.81	0.08
16	1954	99.7	98.3	101.2	100.3	99	101.6	−0.63	−0.26	−0.19	−0.34	−0.05	0.02
17	1778	98.8	97.4	100.2	100.6	99.2	101.9	−1.81	−1.04	−0.76	−1.06	−0.46	0.07
18	649	95.7	93.6	97.8	99.1	97.3	101	−3.41	−2.25	−1.64	−2.11	−1.18	0.15
IQ group[§]													
≤70	949	61.8	60.4	63.2	68.9	67.9	69.9	−6.94	−6.64	−4.86	−4.93	−4.78	0.44
71–75	316	72.9	72.6	73.2	76.1	75.7	76.6	−3.42	−3.22	−2.35	−2.48	−2.27	0.22
76–79	325	77.4	77.1	77.6	79.3	79	79.7	−2.02	−1.61	−1.18	−1.29	−1.07	0.11
80–89	1166	84.8	84.5	85.1	85.5	85.2	85.8	−0.56	−0.48	−0.35	−0.41	−0.3	0.03
90–99	1961	95.1	94.8	95.3	94.5	94.2	94.8	0.70	0.76	0.56	0.52	0.6	0.05
100–109	2584	104.5	104.4	104.7	104	103.7	104.3	0.62	0.64	0.47	0.43	0.5	0.04
110–119	1845	113.6	113.4	113.8	112.8	112.5	113.2	0.81	0.61	0.45	0.41	0.49	0.04
120–124	408	121.9	121.7	122.1	120	119.5	120.5	1.84	1.58	1.16	1.08	1.24	0.11
125–129	254	126.6	126.4	126.7	123.9	123.5	124.4	2.76	2.24	1.64	1.52	1.75	0.15
130+	240	134.1	133	135.3	128.2	127.1	129.2	6.13	4.82	3.53	3.41	3.64	0.32
Sex													
Women	5138	100.4	99.4	101.3	100.3	99.4	101.2	0.15	0.12	0.09	0.04	0.14	0.01
Men	4910	100.4	99.3	101.5	100.2	99.2	101.3	0.10	0.11	0.08	0.03	0.13	0.01
Parent education													
< HS	1667	95.6	94.4	96.8	95.7	94.6	96.7	−0.13	−0.25	−0.18	−0.27	−0.09	0.02
HS graduate	3059	97.6	96.7	98.4	97.8	97	98.5	−0.13	−0.09	−0.06	−0.12	0	0.01
Some college	1988	102.1	101.3	103	101.9	101.1	102.7	0.19	0.29	0.21	0.14	0.29	0.02
College degree	3334	103.9	102.3	105.5	103.5	102	104.9	0.45	0.38	0.28	0.22	0.34	0.03
Income to poverty ratio													
< 1.5	1704	96.9	95.4	98.3	96.9	95.6	98.1	−0.13	−0.1	−0.08	−0.17	0.01	0.01
≤3	2004	99.5	98.5	100.4	99.3	98.5	100.2	0.08	0.1	0.07	−0.01	0.15	0.01
≤6	3078	101.4	100.3	102.4	101	100.1	102	0.37	0.23	0.17	0.1	0.22	0.02
> 6	3262	101.5	100.1	103	101.6	100.2	102.9	0.05	0.14	0.1	0.05	0.17	0.01
Geographic region													
Northeast	1857	101.4	100.4	102.4	101.1	100.4	101.8	0.24	0.16	0.11	0.03	0.19	0.01
Midwest	2760	102.2	101.3	103	102	101.2	102.8	0.12	0.26	0.19	0.12	0.26	0.017
South	3402	99.2	97.3	101.1	99.1	97.3	100.9	0.15	0	0	−0.06	0.06	0
West	2029	99.7	98.1	101.2	99.8	98.1	101.4	−0.01	0.14	0.1	0.03	0.17	0.009
Urbanicity													
Metro	4481	99.9	98.6	101.1	99.9	98.7	101.1	0.10	0.08	0.06	0.01	0.11	0.005
Other Urban	3281	100.8	99.6	101.9	100.6	99.4	101.8	0.16	0.11	0.08	0.02	0.13	0.007
Rural	2286	101.1	100.2	102	100.7	99.8	101.7	0.14	0.27	0.2	0.11	0.29	0.018
Birth Order													
1	3521	101.4	100.4	102.4	101.1	100.2	102	0.09	0.13	0.1	0.04	0.15	0.009
2	3016	100.7	99.4	101.9	100.6	99.5	101.7	0.27	0.11	0.08	0.02	0.14	0.007
3	1765	99.6	98	101.1	99.6	98.1	101	0.11	0.19	0.14	0.05	0.21	0.013
4	891	97.1	95.2	98.9	97.2	95.6	98.8	0.04	−0.03	−0.02	−0.14	0.11	0.002
5+	855	99.2	97.5	100.8	99.3	97.7	100.8	−0.11	0.1	0.07	−0.05	0.19	0.007
Number of siblings													
0	2134	100.8	99.5	102.1	100.7	99.5	101.8	0.08	0.19	0.14	0.07	0.2	0.013
1	3725	100.3	99	101.6	100.1	99	101.3	0.19	0.07	0.05	−0.01	0.1	0.004
2	2438	101.3	99.8	102.7	101.1	99.7	102.5	0.21	0.2	0.14	0.08	0.22	0.013
3	1033	99.3	97.9	100.6	99.3	98	100.6	0.02	0.04	0.03	−0.09	0.14	0.002
4	367	97	93.2	100.8	97.1	93.8	100.4	−0.07	−0.19	−0.14	−0.33	0.06	0.013
5+	285	97.5	94.5	100.4	97.5	94.8	100.1	−0.37	0.37	0.27	0.03	0.51	0.024
Mother age when child was born[¶]													
< 20	1239	99.1	96.8	101.3	98.9	96.9	101	−0.13	0.28	0.21	0.07	0.34	0.019
20–24	2480	100.9	99.5	102.4	100.8	99.5	102.1	0.11	0.31	0.22	0.14	0.31	0.02
25–29	2625	103.4	101.7	105.1	102.9	101.4	104.4	0.57	0.38	0.27	0.2	0.36	0.025
30–34	1787	104.3	102.9	105.7	103.8	102.5	105.1	0.66	0.39	0.29	0.19	0.38	0.026
35+	233	103.9	100.2	107.6	103.4	100.2	106.6	0.59	0.35	0.25	0.03	0.5	0.023
Father age when child was born[¶]													
< 20	522	97.9	94.6	101.2	98	95.2	100.9	−0.09	0.08	0.06	−0.17	0.24	0.006
20–24	1549	101.5	100.1	103	101.3	99.9	102.7	0.15	0.59	0.43	0.32	0.55	0.039
25–29	2359	103.2	101.6	104.7	102.6	101.3	103.9	0.46	0.42	0.31	0.22	0.38	0.028
30–34	1970	104.5	103.1	105.9	104	102.7	105.4	0.56	0.41	0.3	0.24	0.42	0.028
35–39	957	102.9	100.6	105.3	102.6	100.5	104.7	0.63	0.26	0.19	0.05	0.3	0.018
40+	478	104.1	101.6	106.5	103.3	101.1	105.5	0.80	0.39	0.28	0.12	0.48	0.026

[*] Adjusted for age, sex, race/ethnicity, parent education, poverty level, birth order, number of siblings, parent nationality, parent income level, maternal/paternal age when the respondent was born, any psychiatric disorder, and any substance use disorder.

[†] Mean K-BIT difference per decade.

[§] Based on 1989 norms.

[‡] Calculated as the adjusted mean difference divided by the total sample standard deviation (SD = 15).
[¶] Item included those with non-missing parent questionnaire data

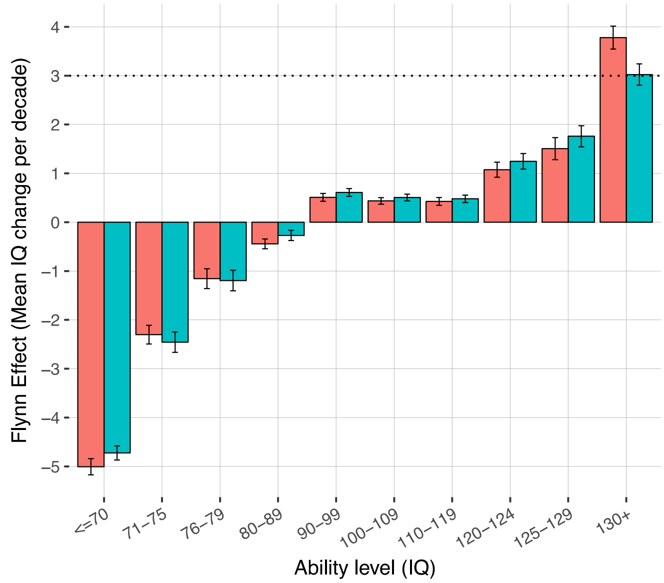

Fig. 1. Flynn effects (Adjusted mean IQ change per decade) by ability level and sex, in a nationally-representative sample of US adolescents. Note: Red = female; Green = male.

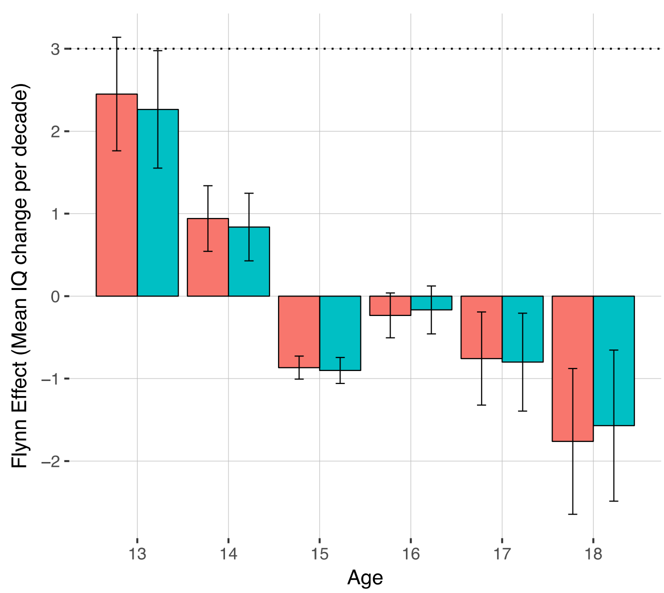

Fig. 2. Flynn effects (Adjusted mean IQ change per decade) by age and sex, in a nationally-representative sample of US adolescents. Note: Red = female; Green = male.

in more recent decades (Pietschnig & Voracek, 2015). However, studies using other tests (e.g., Wechsler scales) did find substantial Flynn Effects (Pietschnig & Voracek, 2015; Trahan et al., 2014). Explanations for the Flynn Effect are diverse. Although genetic explanations focusing on factors such as hybrid vigor (Mingroni, 2007; Rodgers & Wänström, 2007) have been proposed, environmental explanations predominate (Dickens & Flynn, 2001), emphasizing societal changes in perinatal nutrition (Lynn, 2009) and nutrition in general (Colom, Lluis-Font, & Andrés-Pueyo, 2005), education (Teasdale & Owen, 2005), reduced number of siblings (Sundet, Borren, & Tambs, 2008), the prevalence of parasites and the burden of disease (Daniele & Ostuni, 2013; Eppig, Fincher, & Thornhill, 2010), and increased environmental complexity

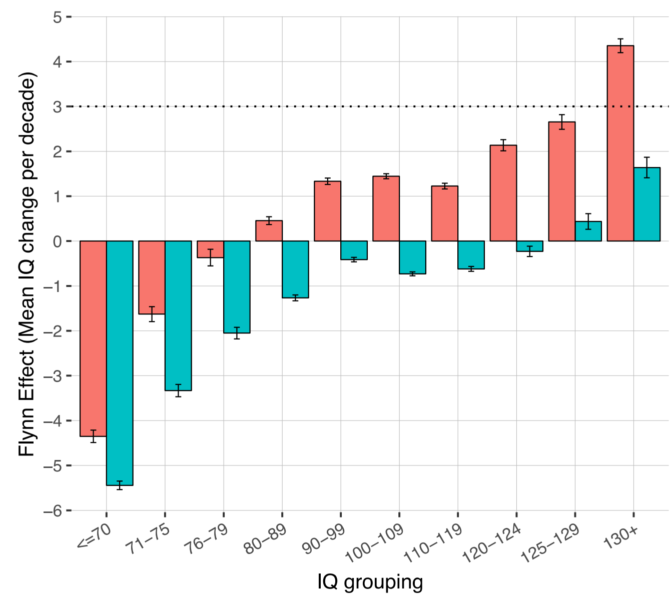

Fig. 3. Flynn effects (Adjusted mean IQ change per decade) by age and ability level, in a nationally-representative sample of US adolescents. Note: Red = younger age (13–15); Blue = older age (16–18).

(Schooler, 1998).

By contrast, other studies have reported reverse Flynn Effects. In discussing these negative trends in Scandinavian countries, Lynn and colleagues hypothesized that they may be due to greater fertility among low SES groups, immigrants, and older adults (Dutton et al., 2016; Dutton & Lynn, 2013). However, a recent analysis in Norway to test these claims largely rejects their hypotheses, reporting that Flynn Effects were not consistent within families over time (Bratsberg & Rogeberg, 2018). Further, a recent meta-analysis found no substantial role of fertility on test score changes across an array of studies (Pietschnig & Voracek, 2015), and recent empirical evidence suggests that immigration effects do not play a meaningful role in explaining Flynn Effect reversals (Pietschnig, Voracek, & Gittler, 2018).

We add to the evidence reported in previous studies, by reporting heterogeneity in the Flynn Effect by ability level and age. We find support for a reverse Flynn Effect for those of low ability and older age, and a positive Flynn Effect for those of high ability and younger age. These results have several implications. First, they signal a widening disparity in the US in terms of cognitive ability, with those at the lower end of the ability dimension not only exhibiting less gains than those at the higher ends, but reversing direction entirely. Second, these results have implications for considering demographic differences when adjusting IQ test scores in the population.

Improvements in education, nutrition, prenatal and post-natal care, and overall environmental complexity over the past century are thought to contribute to the Flynn Effect in the overall population (Dickens & Flynn, 2001; Lynn, 2009; Schooler, 1998; Teasdale & Owen, 2005). However, the disparities by ability level that we identified suggest that the benefits from these societal improvements have been more dramatic for those at the highest ability levels, potentially because they are better able to take advantage of these societal changes. This interpretation is in line with Fundamental Cause Theory (Phelan, Link, & Tehranifar, 2010), which argues that when new knowledge or technology is introduced into a society, those with the highest status are most likely to take advantage first and benefit. Disproportionate utilization by those with higher abilities may widen intellectual disparities, leaving those at the lowest ability levels worse off than before. We note, however, that the Flynn Effect did not differ across other measures of status, such as poverty and parental education. The correlation analyses we conducted revealed a positive association of moderate magnitude between IQ and the size of Flynn Effect, for every age group between 13 and 18, regardless of whether that group showed an overall positive or negative Flynn Effect. One possible interpretation of this pattern is that adolescents with high fluid intelligence, not necessarily those with the highest access to resources, have benefitted most from societal progress over time.

Previous research on the stability of the Flynn Effect across ability levels has produced inconsistent and inconclusive results (McGrew, 2015; Weiss, 2010). Sometimes it has been higher at low IQs, and sometimes a reverse Flynn Effect has been found in high IQ samples (Spitz, 1989; Teasdale & Owen, 1989; Zhou et al., 2010). A meta-analysis examining ability level as a moderator variable did not observe a Flynn Effect for those with low IQ (Trahan et al., 2014). However, previous studies differ in quality (Trahan et al., 2014) and often rely on small sample sizes at the lower end of the IQ distribution (Zhou et al., 2010). Specifically, Trahan and colleagues noted, "the distribution of Flynn effects that we observed at lower ability levels might be the result of artifacts found in studies of groups within this range of ability" (p. 1349).

We also identified variation in the Flynn Effect by age. The positive Flynn Effect of 2.3 points per decade at age 13 approximately equals the value obtained in a summary of studies of Raven's matrices for nearly 250,000 children in 45 countries (Brouwers, Van de Vijver, & Van Hemert, 2009) and in a meta-analysis of about 14,000 children and adults in the US and UK (Trahan et al., 2014). However, the 2-point value is smaller than the traditional 3 points for global intelligence and 4 points for fluid intelligence (Pietschnig & Voracek, 2015). Likewise, the reverse Flynn Effect that occurred at ages 15–18 was similar to effects reported in Scandinavian countries among young adult males during the same time period (Bratsberg & Rogeberg, 2018; Dutton & Lynn, 2013; Sundet et al., 2004; Teasdale & Owen, 2005, 2008), and in other countries as well, such as France (adults tested on WAIS-III and WAIS-IV) and Estonia (young adults tested on Raven's Matrices) (Dutton et al., 2016). The age effects are discordant with previous meta-analyses. Pietschnig and Voracek (2015) evaluated age effects and found stronger gains for adults than children. In their meta-analysis, Trahan et al. (2014) did not find a significant relationship between Flynn Effect and age in their examination of the mean ages across heterogeneous and often small samples. Our methodology differed from the techniques used in both meta-analyses, as we studied large samples that were homogeneous by age.

The notable differences we identify among narrowly defined age groups may be related to cognitive and neurodevelopmental changes that occur during adolescence. Fluid reasoning abilities and cognitive abilities that support reasoning (e.g., rule representation) develop rapidly during early adolescence (Crone et al., 2009; Crone, Donohue, Honomichl, Wendelken, & Bunge, 2006; Ferrer, O'Hare, & Bunge, 2009; Žebec, Demetriou, & Kotrla-Topić, 2015). Brain regions that play a central role in reasoning and problem solving, including the dorsolateral and ventrolateral prefrontal cortex and superior and inferior parietal cortex, also exhibit dramatic changes in structure and function across adolescence (Bunge, Wendelken, Badre, & Wagner, 2004; Ferrer et al., 2009; Gogtay et al., 2004; Wendelken, Ferrer, Whitaker, & Bunge, 2015; Wright, Matlen, Baym, Ferrer, & Bunge, 2008). The notably different Flynn Effects by age in our study caution against generalizing findings for a specific sub-group (such as conscripted young adult males, which comprise the Scandinavian samples) to the nation as a whole (Dutton & Lynn, 2013).

The present study identified no meaningful relationship between Flynn Effect and poverty, parental education other sociodemographic variables and background factors, including parental nationality, birth order, family size, age of birth mother and father. This finding is notable given that these demographic variables are associated with IQ level (von Stumm & Plomin, 2015), including in our sample (Platt, Keyes, et al., 2018).

The results of this study should be considered in light of several limitations. First, the study data were obtained 15 years ago. However, this period was an ideal time to evaluate the presence of a reverse Flynn Effect in the US, given the reverse effects found in Denmark, Norway, Finland, and several other countries (Dutton et al., 2016; Teasdale & Owen, 2008). In more recent years, no reverse Flynn Effect has been observed for Wechsler's scales, as gains on the WAIS-IV (Wechsler, 2008) and WISC-V (Wechsler, 2014). Full Scale IQ have been close to the hypothesized value of 3 points per decade (J Grégoire & Weiss, 2019; Jacques Grégoire, Daniel, Llorente, & Weiss, 2016 Weiss, Gregoire, & Zhu, 2016; Zhou et al., 2010), especially when test content is held constant (J Grégoire & Weiss, 2019; Weiss et al., 2016).

Second, the K-BIT nonverbal test is a screening test that measures a single cognitive ability. It is, however, an analog of Raven's popular matrices test which is commonly used in Flynn Effect studies (Brouwers et al., 2009; Flynn, 1998; Pietschnig & Voracek, 2015). The Flynn Effect is known to differ for different cognitive abilities (e.g., fluid intelligence, short-term memory) (Pietschnig & Voracek, 2015; Teasdale & Owen, 2008), which may contribute to heterogeneity in findings across studies with differing IQ measures. However, the K-BIT and KBIT-2 nonverbal IQ is substantially correlated with comprehensive IQ tests, such as the Wechsler's Full Scale IQ (mid-.50s to mid-70s) (Canivez et al., 2005; Kaufman & Kaufman, 1990, 2004), though it is lower than the correlation between different comprehensive test batteries (Kaufman, 2009; Wechsler, 2014). The present findings are descriptive and any practical application regarding the adjustment of IQs must be made with the awareness that clinical diagnosis, such as the identification of individuals with intellectual disabilities, *must* be based on comprehensive IQ tests such as Wechsler's scales or the Woodcock-Johnson, which assess multiple cognitive abilities.

Third, the study included only adolescents, which represents a narrow period that may not capture meaningful developmental changes. Indeed, fluid reasoning changes between ages 13–18 are minimal (Wechsler, 2008, 2014), including in the present 2003 K-BIT norms sample (Keyes et al., 2016) and the original 1989 norms sample Kaufman & Kaufman (1990, Table 4.7). This age pattern may partially explain why we found no overall Flynn Effect in this sample.

Fourth, different procedures were used to develop the 1989 and 2003 norms. The 1989 norms were estimated based on aggregated data across all age groups, in order to stabilize norms at all ages (Angoff & Robertson, 1987). Although slightly different statistical techniques were used to develop the 2003 norms, the general approach to norms development was similar between samples, and one test author (ASK) was involved in the development of both sets of norms. Both samples were representative of the US distributions of sociodemographic, economic, and other key background variables at the time (Kaufman & Kaufman, 1990; Kessler, Avenevoli, Costello, et al., 2009). Further, both sets of norms are based on six-month age bands. These samples are at least as convergent as similar studies comparing samples used to develop original vs. revised norms. Previous studies have differed substantially by key sociodemographic distributions, such as the WISC and WISC-R (Wechsler, 1949, 1974), which were key samples in the development of the Flynn Effect theory (Flynn, 1984). In the present study, we adjusted the Flynn Effect for an array of background variables to further minimize any differences between the 1989 and 2003 norms samples that may confound the Flynn Effect estimates.

Fifth, the Flynn Effect has had a non-linear trajectory over the past century (Pietschnig & Voracek, 2015). Because our study included IQ measurements at only two time points, we were not able to test the linearity of change over time.

This study is strengthened by the use of a large and representative adolescent sample, with IQs measured with reasoning items that are widely accepted as prototypical measures of fluid intelligence (Dutton et al., 2016). The use of two sets of norms based on a single administration of a test avoids practice effects and bias that may arise from use of different versions of a test.

In conclusion, this study reports important heterogeneity in the Flynn Effect among a nationally-representative sample of US adolescents. We confirmed previous reports of reverse Flynn Effects among large samples of older adolescent males, and extended the same pattern to females. We also found important differential Flynn Effects by ability level. These results add to a growing body of evidence suggesting that Flynn Effect findings from narrow age bands or ability levels may produce divergent findings that do not generalize to the overall population. However, given the potential life or death implications of this research in determining intellectual status in capital punishment cases, the strength of evidence needed for definitive conclusions is extremely high. At this time, we do not have sufficient evidence to recommend differential adjustments to IQ scores. Additional research is needed to replicate the current findings on the full age range and across comprehensive measures of intelligence.

Funding

This research was supported by the National Institute of Mental Health (T32-MH1304346 to Platt; R01-MH103291 and R01-MH106482 to McLaughlin), a Jacobs Foundation Early Career Research Fellowship (McLaughlin), and a OneMind Rising Star Award (McLaughlin).

Disclosures

Alan S. Kaufman is first author of the K-BIT. Although that test is no longer published, he receives royalties on its revision, the KBIT-2.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.intell.2019.101385.

References

Angoff, W. H., & Robertson, G. J. (1987). A procedure for standardizing individually administered tests, normed by age or grade level. *Applied Psychological Measurement, 11*(1), 33–46.

Bain, S. K., & Jaspers, K. E. (2010). Test review: Review of Kaufman brief intelligence test, second edition. *Journal of Psychoeducational Assessment, 28*(2), 167–174. https://doi.org/10.1177/0734282909348217.

Bratsberg, B., & Rogeberg, O. (2018). Flynn effect and its reversal are both environmentally caused. *Proceedings of the National Academy of Sciences of the United States of America, 115*(26), 6674–6678. https://doi.org/10.1073/pnas.1718793115.

Brouwers, S. A., Van de Vijver, F. J. R., & Van Hemert, D. A. (2009). Variation in Raven's Progressive Matrices scores across time and place. *Learning and Individual Differences, 19*(3), 330–338. https://doi.org/10.1016/j.lindif.2008.10.006.

Bunge, S. A., Wendelken, C., Badre, D., & Wagner, A. D. (2004). Analogical reasoning and prefrontal cortex: Evidence for separable retrieval and integration mechanisms. *Cerebral Cortex, 15*(3), 239–249.

Canivez, G. L., Neitzel, R., & Martin, B. E. (2005). Construct Validity of the Kaufman brief intelligence test, Wechsler intelligence scale for children-and adjustment scales for children and adolescents. *Journal of Psychoeducational Assessment, 23*(1), 15–34.

Colom, R., Lluis-Font, J. M., & Andrés-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*(1), 83–91. https://doi.org/10.1016/j.intell.2004.07.010.

Crone, E. A., Donohue, S. E., Honomichl, R., Wendelken, C., & Bunge, S. A. (2006). Brain regions mediating flexible rule use during development. *Journal of Neuroscience, 26*(43), 11239–11247.

Crone, E. A., Wendelken, C., Van Leijenhorst, L., Honomichl, R. D., Christoff, K., & Bunge, S. A. (2009). Neurocognitive development of relational reasoning. *Developmental Science, 12*(1), 55–66.

Daniele, V., & Ostuni, N. (2013). The burden of disease and the IQ of nations. *Learning and Individual Differences, 28*, 109–118.

Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review, 108*(2), 346–369.

Doppelt, J. E., & Kaufman, A. S. (1977). Estimation of the differences between WISC-R and WISC IQS. *Educational and Psychological Measurement, 37*(2), 417–424. https://doi.org/10.1177/001316447703700216.

Dutton, E., & Lynn, R. (2013). A negative Flynn effect in Finland, 1997–2009. *Intelligence, 41*(6), 817–820. https://doi.org/10.1016/j.intell.2013.05.008.

Dutton, E., van der Linden, D., & Lynn, R. (2016). The negative Flynn Effect: A systematic literature review. *Intelligence, 59*, 163–169. https://doi.org/10.1016/j.intell.2016.10.002.

Eppig, C., Fincher, C. L., & Thornhill, R. (2010). Parasite prevalence and the worldwide distribution of cognitive ability. *Proceedings of the Royal Society B: Biological Sciences, 277*(1701), 3801–3808.

Ferrer, E., O'Hare, E. D., & Bunge, S. A. (2009). Fluid reasoning and the developing brain. *Frontiers in Neuroscience, 3*, 3.

Floyd, R. G., Reynolds, M. R., Farmer, R. L., & Kranzler, J. H. (2013). Are the general factors from different child and adolescent intelligence tests the same? Results from a five-sample, six-test analysis. *School Psychology Review, 42*(4), 383.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95*(1), 29–51. https://doi.org/10.1037/0033-2909.95.1.29.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*(2), 171–191. https://doi.org/10.1037/0033-2909.101.2.171.

Flynn, J. R. (1998). Israeli military IQ tests: Gender differences small; IQ gains large. *Journal of Biosocial Science, 30*(4), 541–553. https://doi.org/10.1017/S0021932098005410.

Flynn, J. R. (2012). *Are we getting smarter?: Rising IQ in the twenty-first century.* Cambridge, UK: Cambridge University Press.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., ... Toga, A. W. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences, 101*(21), 8174–8179.

Grégoire, J., & Weiss, L. G. (2019). The Flynn effect and its clinical implications. *WISC-V assessment and interpretation: Scientist-practitioner perspectives* (pp. 245–270). (2nd ed.). San Diego, CA: Academic Press.

Grégoire, J., Daniel, M., Llorente, A. M., & Weiss, L. G. (2016). *The Flynn effect and its clinical implications*.

Homack, S. R., & Reynolds, C. R. (2007). *Essentials of assessment with brief intelligence tests. Vol. 41*. John Wiley & Sons.

Kanaya, T., Scullin, M. H., & Ceci, S. J. (2003). The Flynn effect and U.S. policies: The impact of rising IQ scores on American society via mental retardation diagnoses. *The American Psychologist, 58*(10), 778–790. https://doi.org/10.1037/0003-066X.58.10.778.

Kaufman, A. S. (2009). *IQ testing 101.* Springer Publishing Company.

Kaufman, A. S., Johnson, C. K., & Liu, X. (2008). A CHC theory-based analysis of age differences on cognitive abilities and academic skills at ages 22 to 90 years. *Journal of Psychoeducational Assessment, 26*(4), 350–381.

Kaufman, A. S., & Kaufman, N. L. (1990). *K-BIT: Kaufman brief intelligence test manual.* Circle Pines, MN, US: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004). *KBIT2: Kaufman brief intelligence test* (2nd ed.). Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Wang, J.-J. (1992). Gender, race, and education differences on the K-BIT at ages 4 to 90 years. *Journal of Psychoeducational Assessment, 10*(3), 219–229.

Kessler, R. C., Avenevoli, S., Costello, E. J., Green, J. G., Gruber, M. J., Heeringa, S., ... Zaslavsky, A. M. (2009). Design and field procedures in the US National Comorbidity Survey Replication Adolescent Supplement (NCS-A). *International Journal of Methods in Psychiatric Research, 18*(2), 69–83.

Kessler, R. C., Avenevoli, S., Green, J., Gruber, M. J., Guyer, M., He, Y., ... Zaslavsky, A. M. (2009). National comorbidity survey replication adolescent supplement (NCS-A): III. Concordance of DSM-IV/CIDI diagnoses with clinical reassessments. *Journal of the American Academy of Child & Adolescent Psychiatry, 48*(4), 386–399.

Keyes, K. M., Platt, J., Kaufman, A. S., & McLaughlin, K. A. (2016). Association of fluid intelligence and psychiatric disorders in a population-representative sample of us adolescents. *JAMA Psychiatry, 74*(2), 179–188.

Lynn, R. (2009). What has caused the Flynn effect? Secular increases in the Development Quotients of infants. *Intelligence, 37*(1), 16–24. https://doi.org/10.1016/j.intell.2008.07.008.

McGrew, K. (2015). Norm obsolescence: The Flynn effect. *The Death Penalty and Intellectual Disability*, 155–169.

Mingroni, M. A. (2007). Resolving the IQ paradox: Heterosis as a cause of the Flynn effect and other trends. *Psychological Review, 114*(3), 806.

Phelan, J. C., Link, B. G., & Tehranifar, P. (2010). Social conditions as fundamental causes of health inequalities: Theory, evidence, and policy implications. *Journal of Health and Social Behavior, 51*(1_suppl), S28–S40. https://doi.org/10.1177/0022146510383498.

Pietschnig, J., & Voracek, M. (2015). One century of global IQ gains: A formal meta-analysis of the Flynn effect (1909–2013). *Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 10*(3), 282–306. https://doi.org/10.1177/1745691615577701.

Pietschnig, J., Voracek, M., & Gittler, G. (2018). *Is the Flynn effect related to migration? Meta-analytic evidence for correlates of stagnation and reversal of generational IQ test score changes*.

Platt, J. M., Keyes, K. M., McLaughlin, K. A., & Kaufman, A. S. (2018). Intellectual disability and mental disorders in a US population representative sample of adolescents. *Psychological Medicine*, 1–10.

Platt, J. M., McLaughlin, K. A., Luedtke, A. R., Ahern, J., Kaufman, A. S., & Keyes, K. M. (2018). Targeted estimation of the relationship between childhood adversity and fluid intelligence in a US population sample of adolescents. *American Journal of Epidemiology, 187*(7), 1456–1466.

Raven, J. C. (1936). *Mental tests used in genetic studies: The performance of related individuals on tests mainly educative and mainly reproductive.* Unpublished Master's ThesisUniversity of London.

Reynolds, M. R., Floyd, R. G., & Niileksela, C. R. (2013). How well is psychometric g

indexed by global composites? Evidence from three popular intelligence tests. *Psychological Assessment, 25*(4), 1314.
Rodgers, J. L., & Wänström, L. (2007). Identification of a Flynn Effect in the NLSY: Moving from the center to the boundaries. *Intelligence, 35*(2), 187–196. https://doi.org/10.1016/j.intell.2006.06.002.
Rundquist, E. A. (1936). Intelligence test scores and school marks of high school seniors in 1929 and 1933. *School and Society, 43*(1105), 301.
Salthouse, T. (2010). *Major issues in cognitive aging. Vol. 49*. Oxford University Press.
Sanborn, K. J., Truscott, S. D., Phelps, L., & McDougal, J. L. (2003). Does the Flynn effects differ by IQ level in samples of students classified as learning disabled? *Journal of Psychoeducational Assessment, 21*(2), 145–159. https://doi.org/10.1177/073428290302100203.
Schalock, R. L. (2012). *User's guide: To accompany the 11th edition of intellectual disability: Definition, classification, and systems of supports: Applications for clinicians, educators, organizations providing supports, policymakers, family members and advocates, and health care professionals.* American Association on Intellectual and Developmental Disabilities.
Schooler, C. (1998). Environmental complexity and the Flynn effect. *The rising curve: Long-term gains in IQ and related measures* (pp. 67–79). . https://doi.org/10.1037/10270-002.
Spitz, H. H. (1989). Variations in Wechsler interscale IQ disparities at different levels of IQ. *Intelligence, 13*(2), 157–167. https://doi.org/10.1016/0160-2896(89)90014-7.
von Stumm, S., & Plomin, R. (2015). Socioeconomic status and the growth of intelligence from infancy through adolescence. *Intelligence, 48*, 30–36. https://doi.org/10.1016/j.intell.2014.10.002.
Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The end of the Flynn effect?: A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence, 32*(4), 349–362. https://doi.org/10.1016/j.intell.2004.06.004.
Sundet, J. M., Borren, I., & Tambs, K. (2008). The Flynn effect is partly caused by changing fertility patterns. *Intelligence, 36*(3), 183–191.
Teasdale, & Owen, D. R. (2005). A long-term rise and recent decline in intelligence test performance: The Flynn Effect in reverse. *Personality and Individual Differences, 39*(4), 837–843. https://doi.org/10.1016/j.paid.2005.01.029.
Teasdale, & Owen, D. R. (2008). Secular declines in cognitive test scores: A reversal of the Flynn Effect. *Intelligence, 36*(2), 121–126. https://doi.org/10.1016/j.intell.2007.01.007.
Teasdale, T. W., & Owen, D. R. (1989). Continuing secular increases in intelligence and a stable prevalence of high intelligence levels. *Intelligence, 13*(3), 255–262.
Trahan, L. H., Stuebing, K. K., Fletcher, J. M., & Hiscock, M. (2014). The Flynn effect: A meta-analysis. *Psychological Bulletin, 140*(5), 1332–1360. https://doi.org/10.1037/a0037173.
Wang, J.-J., & Kaufman, A. S. (1993). Changes in fluid and crystallized intelligence across the 20-to 90-year age range on the K-BIT. *Journal of Psychoeducational Assessment, 11*(1), 29–37.
Wechsler, D. (1949). *Wechsler intelligence scale for children*.
Wechsler, D. (1974). *Wechsler intelligence scale for children-revised.* Psychological Corporation.
Wechsler, D. (2008). *Wechsler adult intelligence scale–fourth edition (WAIS–IV). 22*, San Antonio, TX: NCS Pearson498.
Wechsler, D. (2014). *WISC-V: Technical and interpretive manual.* NCS Pearson, Incorporated.
Weiss, L. G. (2010). Considerations on the Flynn effect. *Journal of Psychoeducational Assessment, 28*(5), 482–493. https://doi.org/10.1177/0734282910373572.
Weiss, L. G., Gregoire, J., & Zhu, J. (2016). Flaws in Flynn effect research with the Wechsler scales. *Journal of Psychoeducational Assessment, 34*(5), 411–420. https://doi.org/10.1177/0734282915621222.
Wendelken, C., Ferrer, E., Whitaker, K. J., & Bunge, S. A. (2015). Fronto-parietal network reconfiguration supports the development of reasoning ability. *Cerebral Cortex, 26*(5), 2178–2190.
Wright, S. B., Matlen, B. J., Baym, C. L., Ferrer, E., & Bunge, S. A. (2008). Neural correlates of fluid reasoning in children and adults. *Frontiers in Human Neuroscience, 2*, 8.
Žebec, M. S., Demetriou, A., & Kotrla-Topić, M. (2015). Changing expressions of general intelligence in development: A 2-wave longitudinal study from 7 to 18 years of age. *Intelligence, 49*, 94–109. https://doi.org/10.1016/j.intell.2015.01.004.
Zhou, X., Gregoire, J., & Zhu, J. (2010). The Flynn effect and the Wechsler scales. In L. G. Weiss, D. H. Saklofske, D. Coalson, & S. E. Raiford (Eds.). *WAIS-IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 141–166). (1 ed.). London: Academic Press.