Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Personality and Individual Differences 39 (2005) 837–843
PERSONALITY AND INDIVIDUAL DIFFERENCES
www.elsevier.com/locate/paid

A long-term rise and recent decline in intelligence test performance: The Flynn Effect in reverse

Thomas W. Teasdale [a,*], David R. Owen [b]

[a] *Department of Psychology, University of Copenhagen, Njalsgade 88, 2300 Copenhagen S, Denmark*
[b] *Department of Psychology, Brooklyn College, City University of New York, NY 11210, United States*

Received 13 September 2004; accepted 17 January 2005
Available online 23 May 2005

Abstract

In the 1980s reviewed evidence indicated that, through the preceding decades of the last century, population performance on intelligence tests had been rising substantially, typically about 3–5 IQ points per decade, in developed countries. The phenomenon, now termed the 'Flynn Effect', has been variously attributed to biological and/or to social and educational factors. Although there is some evidence to suggest a slowing of the effect through the 1990s, only little evidence, to our knowledge, has yet been presented to show an arrest or reversal of the trend. Substantially replicating a recent report from Norway, we here report intelligence test results from over 500,000 young Danish men, tested between 1959 and 2004, showing that performance peaked in the late 1990s, and has since declined moderately to pre-1991 levels. A contributing factor in this recent fall could be a simultaneous decline in proportions of students entering 3-year advanced-level school programs for 16–18 year olds.

Keywords: Flynn effect; Intelligence; Denmark

* Corresponding author. Tel.: +45 35 32 87 15; fax: +45 35 32 87 45.
E-mail address: tom.teasdale@psy.ku.dk (T.W. Teasdale).

0191-8869/$ - see front matter
doi:10.1016/j.paid.2005.01.029

1. Introduction

It is now over 20 years since Flynn (1984), in a seminal review, first drew attention to the extensive evidence for rising levels of intelligence test scores in the American population through the preceding decades of the last century. This was followed by a further review in which he demonstrated the same effect to have occurred in other economically developed countries from which relevant evidence was available (Flynn, 1987). The magnitude of the effect, which has come to be dubbed the *Flynn Effect* (Herrnstein & Murray, 1996), varied in time and place but could generally be summarized to be about 3–5 IQ points per decade. The effect has been seen most prominently in so-called 'fluid' tests of intelligence, i.e., tests requiring educative reasoning to a logical conclusion from given, usually abstract, information such as is presented in Raven's Progressive Matrices (Raven, 2000). The effect has typically been ascribed to either biological factors, such as improved nutrition and health care, or social factors including educational developments (Neisser, 1998).

Although evidence of the Flynn Effect continues to be reported in the literature, such reports are most often based on two time-point comparisons, which preclude the possibility of examining whether or not the gains have been linear over time. Furthermore, there have, to our knowledge, been few reports concerning any continuing Flynn Effect specifically over the time period after the 1980s and into the present century. One exception is a Danish study using draft board data, in which we reported, inter alia, a slowed rate of gain between 1988 and 1998 in Denmark (Teasdale & Owen, 2000). More recently, Sundet, Barlaug, and Torjussen (2004), using 50 years of conscript data in Norway, have reported a complete cessation of gains between the mid 1990s and 2002.

There has long been a recognition of the unlikelihood of the effect continuing indefinitely. In the present study we update our earlier reports to include the time period up to and including 2004, particularly to investigate whether in Denmark, as in Norway, the Flynn Effect has ended.

2. Method

Our data derive from the Danish draft board. There has been conscription in Denmark uninterruptedly since Word War II and, at age 18, all Danish men are required to appear before the draft board, which assesses their suitability for military service. The only exemptions are about 5–10% of men who have a medically documented illness or condition which would disqualify them from military service, e.g., severe asthma, Scheuermann's disease. The assessment procedure includes an intelligence test, the Børge Priens Prøve (BPP, Rasch, 1980). This is a 45-min paper-and-pencil test administered in groups of about 30. It comprises four subtests, (a) a matrix test resembling Raven's Progressive Matrices (19 items, 15 min), (b) a verbal analogy test (24 items, 5 min), (c) a number sequence test (17 items, 5 min), and (d) a geometric figures test (18 items, 10 min) (Teasdale & Owen, 1994). None of the subtests is multiple-choice. For administrative purposes the BPP test is scored as the total number of correct items, 0–78, and this raw score is converted to a stanine scale, 1–9. The total BPP score has been shown to correlate .82 with the Wechsler Adult Intelligence Scale (Mortensen, Reinisch, & Teasdale, 1989).

The BPP test was introduced by the draft board in 1957 and has been used continuously and unaltered, in both content and administration procedure, to the present day. The grouping ranges for the stanine conversions were derived using data from large pilot samples of conscripts in the early 1950s to yield an approximately normal distribution, mean = 5 SD = 2, and likewise continue to be used unaltered.

The draft board also registers educational level. The scale used has varied over time in keeping with changes in the Danish educational system. Since 1991, however, it has consistently recorded whether or not the person had attended, or was currently attending, a 3-year advanced-level college (ALC) or 'gymnasium'. School attendance is compulsory in Denmark until age 16. Beyond this age, the more able school-leavers typically attend such gymnasium colleges which cover an academic curriculum and which culminate in an examination-based certification required to pursue university, or similar, further education. Other school-leavers most commonly enter shorter, less academic, educational programs, often directed towards specific occupational training.

About 25,000–35,000 men are tested annually, the overwhelming majority at ages 18 or 19. The declining numbers since the 1960s reflect a generally declining birth-rate in Denmark. We have obtained (anonymous) results for the BPP stanine scores for 1959 (the latter half of the year), 1969, 1979 and all years from 1989 to 2004, together with the codes for educational level for all years from 1991 to 2004. In total our data comprise 549,149 subjects.

3. Results

Table 1 shows the year-by-year mean scores for the BPP, together with higher order statistics. It can be seen that performance on the test improved at a decelerating rate between 1959 and

Table 1
Descriptive statistics for BPP test as a function of year of testing

Year	Mean	SD	Skewness	% Scores 1 and 2	% Scores 8 and 9	*n*
1959	4.67	2.02	0.05	15.6	8.4	16,618
1969	5.09	1.90	−0.15	9.9	9.4	37,212
1979	5.53	1.77	−0.39	6.1	11.7	36,143
1989	5.76	1.63	−0.45	3.6	12.3	34,123
1990	5.77	1.63	−0.47	3.6	12.3	35,102
1991	5.81	1.61	−0.49	3.3	12.5	34,106
1992	5.85	1.59	−0.47	3.0	13.1	32,920
1993	5.90	1.60	−0.51	3.0	13.8	31,010
1994	5.91	1.60	−0.50	2.9	13.8	31,454
1995	5.91	1.58	−0.47	2.7	13.8	29,505
1996	5.91	1.58	−0.49	2.8	13.6	28,118
1997	5.91	1.56	−0.49	2.8	13.4	28,460
1998	5.94	1.58	−0.49	2.8	14.4	25,282
1999	5.84	1.59	−0.46	3.1	12.8	26,582
2000	5.75	1.63	−0.46	3.9	12.1	25,893
2001	5.78	1.61	−0.45	3.5	12.3	24,797
2002	5.79	1.61	−0.48	3.5	12.4	24,411
2003	5.77	1.60	−0.46	3.6	12.0	23,907
2004	5.77	1.59	−0.47	3.3	11.7	23,505

1998, since which year there has been some decline. The means for all years of this century have been lower than for 1991. It can also be seen that the period of increasing means was marked by decreasing variances and increasingly negative skewness with both of these trends being somewhat reversed since 1998. The reduced variance and increased skewness are attributable to asymmetrical developments in the distribution of test scores. Since 1959 there has been a very marked reduction in the percentage of men obtaining the lowest two stanine scores (1 and 2), contrasting with a much more modest increase in the percentage obtaining the highest two scores (8 and 9) (see Table 1).

The changing variances make an estimation of the magnitude of the gains problematic, but Fig. 1 shows 'IQ' change across the 45-year period standardized against the mean and standard deviation of the stanine scores in 1959. As can be seen, calculated in this way, the gains amount to a little over 3 points for the decades 1959–1969 and 1969–1979, and for 1979–1989 decade the gain approached 2 IQ points. For the 9 years from 1989 through the peak year of 1998, the gain was only about 1.3 points where after all of this last small gain had been lost by 2000 and has not been recovered since. It should be noted that all of these 'IQ' changes would be of greater magnitude if the standard deviation for some year later than 1959 had been used for the standardization.

The modest decline of recent years has been accompanied by similar small changes in educational levels. Fig. 1 also shows attendance at an Advanced Level College (ALC) to have risen from 38.9% in 1991 to 49.1% in 1998, followed by a decline to 44.3% in 2000 and a partial recovery to 48.1% in 2004.

The two functions are not perfectly parallel; the post-1998 decline in test scores reaches a level below that of 1991, whereas the ALC percentage since 1998 has remained well above that of 1991. It is nonetheless suggestive that both should peak in the same year, 1998, and reach their lowest post-1998 values in the same year, 2000. It is also suggestive that the year-by-year changes in mean

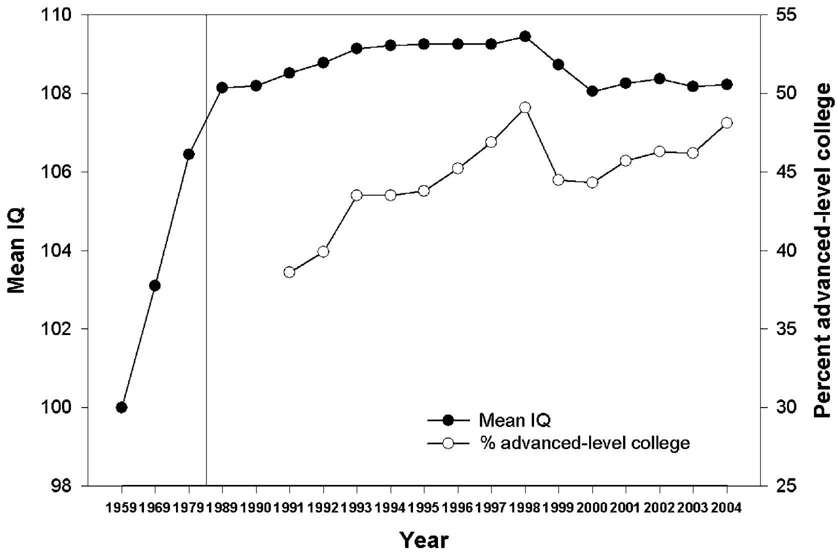

Fig. 1. Mean 'IQ' and percentage of men attending Advanced-Level Colleges as a function of year of testing. For the 1959 half-year cohort ($n = 16{,}618$) the 95% CI of the mean IQ is ±0.23. For all subsequent full-year cohorts ($23{,}505 \leqslant n \leqslant 37{,}213$) it is within the range ±0.13 to ±0.15.

test score and ALC attendance between 1991 and 2004 correlate very highly ($r = .82$, $p < .05$). Furthermore, test scores are strongly related to ALC attendance. The point biserial correlations between test score and ALC attendance all lie within the narrow range between .50 and .53 for each year between 1991 and 2004 and, at each of those years, the mean 'IQ' difference between those attending and those not attending ALCs is 12 points. Taken together this evidence suggests that changes in educational level, in particular the decline in ALC attendance since 1998, might be a contributing causal factor in the modest fall in mean test scores since that year.

4. Discussion

It is a limitation of the present study that, by its nature, it reports on the Flynn Effect in men only. Although the strongest evidence on the issue has always derived from data on conscripts, where the same test is often administered over many years to large and representative numbers of men of very similar ages at testing, this has left open the question of whether a similar effect has occurred among women. However, in the only study of conscripts including both men and women, Flynn (1998) reported on Israeli data showing virtually the same gains for both sexes over the period from 1971 to 1984.

A second limitation of the present study is that we have not been able to report, across the 46 years spanned by our data, on the four individual subtests which comprise the total BPP score. We have earlier reported (Teasdale & Owen, 2000) that almost all of the modest gain between 1988 and 1998 derived from the geometric figures test of spatial ability. Preliminary analysis of data from 2003 and 2004 suggests that all four subtests have shown declines since the 1998 peak.

The test score gains which we have found between the late 1950s and the late 1990s, and which we have already reported elsewhere (Teasdale & Owen, 2000), are smaller than, but comparable to, those originally reviewed by Flynn (1984, 1987). There has been much less evidence concerning the changed shape of the test score distribution which we report, specifically the considerable reduction in proportions with very low scores with only a much smaller increase in the proportions with very high scores. We have considered this to be primarily related to developments in the Danish education system through the early decades following World War II, e.g., pre-school care, earlier school starting and later leaving, and increased resources allocated to children with special educational needs, together with greater pedagogical emphasis on problem-solving approaches to learning rather than rote-learning (Teasdale & Berliner, 1991; Teasdale & Owen, 1989, 1994).

In a recent report, Sundet et al. (2004), using draft board data in Norway between 1957 and 2002, found a very similar secular trend to that which we have found for Denmark for the same time period, namely early gains primarily at the low end of the test score distributions, followed by stagnation through the 1990s. The similarity between the two Scandinavian countries, with close historical, geographical, linguistic and cultural links, is striking.

Sundet et al. (2004) reported on three subtests and noted that the subtests in which the greatest gains had been seen perhaps had also been susceptible to a ceiling effect, perhaps therefore artificially creating the negative skewness seen in both countries. It is unlikely, however, that the disappearance of the Flynn Effect in Denmark results simply from a ceiling effect of the test itself. Unlike the situation with the Norwegian test and, for instance, Raven's Standard Progressive Matrices (Raven, 2000), there has always remained considerable room for gains at high levels to have manifested themselves in the BPP test. The stanine scores of 8 and 9, which, as seen in Table 1, have never been achieved by as much as 15% of all men in any year, correspond to the range of scores greater than 53 of the 78 items in the test.

Other possible artefactual causes for the cessation of the Flynn Effect need also to be considered. The four subtests of the BPP all involve abstract reasoning, or 'fluid' intelligence, and careful examination of the test items has not suggested that the recent decline is attributable to any incipient obsolescence of individual test items, albeit that they were developed some 50 years ago. This issue, however, needs to be examined more closely. Another possibility, voiced widely on the web but as yet with little hard evidence, is that paper-and-pencil tests, such as the BPP, could have become increasingly unfamiliar to generations of school-leavers whose school education has been increasingly dominated by computer-use.

If the recent modest decline in intelligence test performance is real rather than an artefact, then a potential contributor to it is the simultaneous—modest—decline in attendance at advance level colleges. As noted above, we have earlier reported on parallel improvements in the intelligence test scores and improvements in Danish educational levels, as well as the high correlations between them. If other aspects of the Danish educational system are no longer advancing at the rate seen in earlier decades, then a small decline in the percentage of school-leavers entering advanced-level colleges could perhaps account for a similar decline in test performance. The reason for the decline in advanced-level college attendance itself is not easy to identify, but it is possible that a growth of shorter, more practically-oriented post-school educational programs has attracted increasing numbers of school-leavers.

We recognise, however, the impossibility of proving causal relations from correlational data, and the possibility would need to be considered, that some other recent societal changes in Denmark have resulted in modest reductions in cognitive abilities which in turn has made the challenge of advanced level education less attractive to 16-year-old school-leavers.

Whatever the cause of the decline, however, it is striking that the mean test scores through each of the first four years of the present decade have invariably been lower than all means of the last decade after 1991. The Flynn Effect, in Denmark, as in Norway, would appear to have ended. In presenting his original evidence, Flynn (1987) was careful to point out that the eponymous effect could be documented in developed nations only, in which category Norway and Denmark obviously belong. One might therefore speculate that a cessation of the Flynn Effect would now similarly be found in other developed countries. In contrast, there is recent evidence of continuing increases in intelligence in the developing world (Cocodia et al., 2003; Daley, Whaley, Sigman, Espinosa, & Neumann, 2003), perhaps because, unlike in developed countries, there continues to be widespread and substantial improvement in educational levels, although, as Daley et al. (2003) point out, nutritional and health care factors could also play a role. If these two opposing trends continue, then any differentials in cognitive abilities between currently developed and developing countries (Lynn & Vanhanen, 2002) could be expected to diminish in the future.

Acknowledgments

We are much indebted to Stig Meincke of the Military Psychology Unit at The Danish Ministry of Defence for giving us access to the data.

References

Cocodia, E. A., Kim, J. S., Shin, H. S., Kim, J. W., Ee, J., Wee, M. S. W., et al. (2003). Evidence that rising population intelligence is impacting in formal education. *Personality and Individual Differences, 35*, 797–810.

Daley, T. C., Whaley, S. E., Sigman, M. D., Espinosa, M. P., & Neumann, C. (2003). IQ on the rise—The Flynn effect in rural Kenyan children. *Psychological Science, 14*, 215–219.

Flynn, J. R. (1984). 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. (1998). Israeli military IQ tests: Gender differences small. *Journal of Biosocial Science, 30*, 541.

Herrnstein, R. J., & Murray, C. (1996). *The bell curve: Intelligence and class structure in American Society*. New York: Simon & Schuster.

Lynn, R., & Vanhanen, T. (2002). *IQ and the wealth of nations*. New York: Praeger Publishers.

Mortensen, E. L., Reinisch, J. M., & Teasdale, T. W. (1989). Intelligence as measured by the WAIS and a military draft board group test. *Scandinavian Journal of Psychology, 30*, 315–318.

Neisser, U. (1998). *The rising curve: Long-term gains in IQ and related measures*. Washington: American Psychological Association.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.

Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology, 41*, 1–35.

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*, 349–362.

Teasdale, T. W., & Berliner, P. (1991). Kindergarten attendance in relation to educational-level and intelligence in adulthood—a geographical analysis. *Scandinavian Journal of Psychology, 32*, 336–343.

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. (1994). Thirty-year secular trends in the cognitive abilities of Danish male school-leavers at a high educational level. *Scandinavian Journal of Psychology, 35*, 328–335.

Teasdale, T. W., & Owen, D. R. (2000). Forty-year secular trends in cognitive abilities. *Intelligence, 28*, 115–120.