Available online at www.sciencedirect.com
ScienceDirect
Intelligence 36 (2008) 183 – 191
INTELLIGENCE

# The Flynn effect is partly caused by changing fertility patterns

Jon Martin Sundet [a,*], Ingrid Borren [b], Kristian Tambs [a]

[a] National Institute of Public Health, P.O. Box 4404 Nydalen, N-0403 Oslo, Norway
[b] Institute of Psychology, University of Oslo, Norway

Received 15 November 2006; received in revised form 24 April 2007; accepted 24 April 2007
Available online 8 June 2007

---

## Abstract

The present study investigates the impact of changing fertility patterns on the Flynn effect. Intelligence test data comprised scores of army conscripts on an arithmetic, language and a Raven-similar test, and a composite score (General Ability). Family data of the conscripts enabled a decomposition of the population mean into effects of sibship size on the mean intelligence and the proportion of the persons comprising the various sibship sizes within each of 13 birth cohort groups (each comprising 3 birth years). Both the means within each sibship and the proportions of the different sibship sizes varied across cohorts. Estimated changes in means due to changing proportions of sibship sizes alone were calculated by fixing the mean intelligence test score within the different sibship sizes at the level of the oldest birth cohort (1938–1940) and letting the proportions of the different sibship sizes take their empirical values in each of the subsequent 12 three-year cohort groups. It is concluded that changing proportions of sibship sizes had a moderate effect both on General Ability and the subtest scores, and that most of the changes were connected to changing sibship means across cohorts.

*Keywords:* Fertility changes; Flynn effect

---

## 1. Introduction

A secular increase in mean intelligence test scores – the Flynn effect – has been observed in more than 20 industrialized countries (Flynn, 1987, 1999). Recently a substantial Flynn effect was observed in rural Kenya (Daley, Whaley, Sigman, Espinosa & Neumann, 2003). Secular increases seem to be largest in non-verbal tests, like Ravens Progressive Matrices (Flynn, 1987; Sundet, Barlaug & Torjussen, 2004). In Norway, testing of military male conscripts shows that most of the increases in mean intelligence have taken place from cohorts born before the Second World War to the birth cohorts shortly after the war (Sundet et al., 2004). After this period the increase rates have been substantially smaller among conscripts both in Norway and in nearby Denmark (Sundet et al., 2004; Teasdale & Owen, 2000). In these two countries the Flynn effect seems to have come to a complete stop, or even reversed in recent years (Sundet et al., 2004; Teasdale & Owen, 2005).

Both the nature of the secular trends in IQ scores and their causes has been extensively discussed. Rodgers (1998) points out that the Flynn effect could be due to changes within certain ranges of the IQ distribution. Thus, it seems that the Flynn effect in some countries is partly or mostly due to lower prevalence of low scorers in more recent cohorts (Lynn & Hampson, 1986; Sundet et al., 2004; Teasdale & Owen, 1989, 2000). Some studies have reported a declining *g*-factor in more recent cohorts (Kane

---

* Corresponding author.
*E-mail address:* JOSU@fhi.no (J.M. Sundet).

0160-2896/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2007.04.002

& Oakland, 2000; Lynn & Cooper, 1993; Sundet et al., 2004). Wicherts, Dolan, Hessen, Oosterveld, van Baal, Boomsma and Span (2004) analyzed several data sets and found indications of lacking factorial stability across cohorts. These findings certainly complicate the interpretation of the Flynn effect, because changes in mean intelligence test scores across cohorts may partly be due to changes in different underlying constructs.

Doubts about the validity of intelligence test scores as indicators of "real" intelligence have been voiced from time to time (e.g. Flynn, 1987). This is a difficult question, but some studies indicate that the Flynn effect has been accompanied by "real world" indications of rising population intelligence (Howard, 1999, 2001).

Although a lot of potential causes of the Flynn effect, e.g. nutrition, education, and access to mass media have been widely discussed (cf. Neisser, 1998), there are still others to investigate more thoroughly. Rodgers (1998) pointed out that our knowledge about the occurrence or non-occurrence of a Flynn effect in subgroups is insufficient. Zajonc and Mullally (1997) called attention to the potential utility of studying changes of family configurations across generations. This possibility arises because there is a connection between family size and intelligence. Two factors connected to family size might be considered. One of them is the well-documented negative correlation between sibship size and intelligence test scores (Belmont & Marolla, 1973). The other is the alleged effect of birth rank order on intelligence test scores (Zajonc, 1976; Zajonc & Mullally, 1997). The empirical status of the birth order effect on intelligence is at present unclear. Recently, it has been argued that the birth order effect on intelligence is a myth created by inadequate study designs (Rodgers, Cleveland, Van den Oord, & Rowe, 2000). The birth order effect issue may be clarified by future meta-analytic studies addressing important methodological concerns like selection problems, cross-sectional versus longitudinal designs, and sample size. In the present paper we have decided to use sibship size as the main analytical unit.

Two possible factors connected to sibship size may contribute to the Flynn effect; changes of the mean intelligence within sibships across cohorts, and cohort changes of the relative number of persons comprising the various sibship sizes. This can be seen by considering a hypothetical population where the sibship sizes range from singletons to families comprising 5 siblings, with 20% of the persons belonging to each of the sibship sizes at a given time. The IQ means decrease by 2.5 IQ points across sibship sizes, ranging from 105 for singletons to 95 for persons in 5-sibships. If the mean IQ's within each sibship size and the proportions of the different sibship sizes remain constant over time, no Flynn effect will be seen in our hypothetical population (the aggregated population mean, calculated as $105*0.2+102.5*0.2+100*0.2+97.5*0.2+95*0.2$, will remain at 100 IQ points). However, consider that the mean IQ's within sibship sizes remain constant over time, whereas the relative proportions of the different sibship sizes vary over time. Say that the proportion of singletons increases from 20% at time 1 to 40% at time 2, the proportion of persons comprising 2-sibships changes from 20% to 30%, 3-sibships from 20% to 15%, 4-sibships from 20% to 10%, and 5-sibships from 20% to 5%. The population mean will then change from 100 at time 1 to 102.25 at time 2 $(105*0.40+102.5*0.30+100*0.15+97.5*0.10+95*0.05)$. In this hypothetical case, a Flynn effect has been produced by changing proportions of sibship sizes alone, without any changes within the subgroups of different sibship sizes. A similar change could be produced by changing IQ means within each sibship size without any proportion changes. Thus, if the proportions remained constant over all sibships, but the mean increased by 2.25 IQ points in all sibship sizes from time 1 to time 2 the population mean would also be 102.25, i.e. identical to the population mean in the first case.

The interpretations of the increase of the population mean IQ in these two cases would obviously be very different. This reasoning indicates that it might be reasonable to distinguish two broad classes of causes of the Flynn effect: Factors that cause generational changes within subgroups (and the population mean), and on the other hand, population mean changes caused by changing proportions of these subgroups in the population. Actually, it is quite tempting to consider changes in observed means not apparent in subgroups but only due to changing proportion of these subgroups as aggregation artifacts, and that a "real" Flynn effect is the change in observed means after the effect of proportion changes has been removed. It is therefore of considerable interest and importance to clarify the relative importance of the effect of fertility changes and changes of the means within sibship sizes over time.

There are few studies directly concerned with estimating the effects of fertility changes. Zajonc (1976) argued that changing family configuration might affect the population mean scores of cognitive abilities tests. Alwin (1991) investigated the reasons for declining verbal SAT scores in the US, and did not find any evidence indicating effects of changing family configuration. The aim of the present paper is to investigate the relative contribution of fertility changes and increasing IQ means within sibship sizes to the Flynn effect in the intelligence test scores of Norwegian conscripts (mostly males) in the birth cohorts from 1938 to 1985.

## 2. Materials and methods

In Norway, military service is compulsory for every able young man. Before they enter the service the young men are required to meet before a draft board where their medical and psychological suitability, including intellectual ability, for military service are assessed. A great majority of the men meeting before the draft board (about 95%) are examined between their 18th and 20th birthdays. Physically or psychologically disabled are exempted from these investigations. Also, seamen and others being abroad at the normal conscript age are normally exempted. Females may meet on a voluntary basis.

### *2.1. Test materials*

Intelligence test data comprised scores on three speeded subtests (Arithmetic, Word Similarities and Figures) and a composite score (General Ability) in stanine units ($M$=5, SD=2). General Ability scores were obtained by transforming the raw scores on each subtest in a standardization sample into normally distributed $F$-scores, and subsequently transformed into stanine scores. The Arithmetic test, presented in prose, purports to measure arithmetic and elementary algebraic ability but also logical reasoning ability, and is quite similar to the Arithmetic test in the Wechsler Adult Intelligence Scale (WAIS). The Word Similarities test (similar to the Vocabulary test in WAIS) is a synonym test. The Figures test was constructed to be very similar to Raven Progressive Matrices. The test–retest reliabilities of Arithmetic, Figures and Word Similarities as calculated from a sample ($N \approx 800$) in the mid 1950's were .84, .72 and .90, respectively. Alpha coefficients of Arithmetic, Figures and Word Similarities calculated for the draft cohorts 1993–2002 were .81, .80 and .90, respectively (Sundet et al., 2004). In a smallish sample ($N$=48), the correlation between General Ability and the WAIS IQ has been found to be .73, with time spans between the two tests varying from two to 25 years (Sundet, Tambs, Harris, Magnus & Torjussen, 2005).

The contents of the Figures and Word Similarities tests were unchanged over the period of the present study. The Arithmetic test was slightly modified in 1963. This change was mainly a modernization of some of the items.

The intercorrelations between the tests were in the .50's to .60's. The factor loadings were quite similar, but largest for the Word Similarities test and smallest for the Figures test.

### *2.2. Scales and norms*

Intelligence data partly comprised General Ability scores in stanine units and partly subtest scores in raw score units and/or scaled scores. The norms for calculating the General Ability scores were originally constructed in 1954. They were changed in 1963, and these norms were identical to the 1954 norms except that they were stricter for the Arithmetic subtest. Sundet et al. (2004) estimated that this change corresponded to around 6.3 IQ points on the Arithmetic subtest. The whole battery was renormed in the mid 1970's and the new norms were fully implemented in 1980. In the present paper we have scaled all the General Ability scores according to the norms from 1954. General Ability scores calculated according to the 1963 norms have been elevated by 2.1 IQ units (6.3/3), and 9.6 IQ points have been added to the General Ability scores calculated according to the 1980 norms (2.1+7.5).

Raw scores on the Figures and Word Similarities subtests were scaled directly into stanine scores according to the norms from 1954. Scaled scores on the Arithmetic test were obtained by adding 6.3 IQ unit equivalents to the scores calculated from the 1963 norms for this test. In two cases (the 1958 and 1959 draft cohorts) the scores were given in accordance with norms from 1961 that have been unavailable to us. In the 1957 draft cohort both raw scores and scaled scores according to the 1961 norms were available. Comparing the scaled scores calculated from the raw scores according to the 1954/1961 norms with the scores given according to the 1961 norms, it was possible to estimate the 1954 score equivalents in the data sets from the draft cohorts in 1958 and 1959. All the intelligence test scores have been transformed into the more common IQ units relative to $M$=100 and SD=15 according to the norms from 1954. (A more complete description of the scaling procedures may be found in Sundet et al., 2004).

### *2.3. Subjects and data*

General Ability data in stanine units were retrieved for the draft cohorts 1969–2003 (≈birth cohorts 1950–1985). Subtest data were available for a substantial subset of the draft cohorts 1993–2003. Data from the draft cohorts 1957–1959 comprised both subtest data and General Ability scores.

Data on sibship size (defined as the proportion of persons in a given sibship size) were partly attained from the draftee (1938–1940 birth cohorts), and partly from the governmentally held Statistics Norway family registers. The final data file comprised 992,274 persons

Table 1
Frequencies and proportions (%) of persons by sibship size

| Sibship size | *N* | % |
|---|---|---|
| 1 | 67,798 | 6.8 |
| 2 | 329,719 | 33.2 |
| 3 | 320,296 | 32.3 |
| 4 | 158,891 | 16.0 |
| 5 | 64,282 | 6.5 |
| 6+ | 51,288 | 5.2 |
| Total | 992,274 | |

having General Ability scores and sibship size data. 7902 (0.7%) were females, mainly from comparatively recent cohorts. The corresponding numbers for the Arithmetic, Figures and Word Similarities were 211,576, 211,074 and 211,206 persons. Sibship size data were ascertained for over 99% of those having General Ability data.

Sibship size ranged from 1 to 18. Sibships larger than 6 have been included in the subgroup comprising 6 siblings. We have studied General Ability means in sibship sizes from 1 to 6+, and relative proportions of sibship sizes in 13 cohort groups, each comprising three years. Subtest means were studied in two cohort groups (1938–1940 and 1974–1985).

Table 1 displays the frequencies of persons in the different sibship sizes for persons with General Ability scores.

### 2.4. Analysis strategy

The mean intelligence test score in the birth cohort $j$ comprising $k$ sibship sizes may be written as $M_j=\Sigma M_{ji}*p_{ji}$ ($i$ runs from 1 through $k$), where $M_j$ is the intelligence mean in cohort $j$, $M_{ji}$ is the mean intelligence of the $i$th sibship size in cohort $j$, and $p_{ji}$ is the proportion of the $i$th sibship size in cohort $j$. According to this algebraic identity the mean of each birth cohort group has been decomposed into the mean intelligence test score within each sibship size and the proportion of each sibship size in each cohort group. Cohort changes in mean intelligence test scores within each sibship size, and changes in the proportion of each sibship size across cohorts may thus be studied separately. In addition, it is possible to study the course of the Flynn effect in various hypothetical, but theoretically interesting scenarios. We may imagine that the mean intelligence scores within sibships remained constant over all cohort groups, whereas the proportions of the different sibships across cohorts varied just as observed (this is the scenario illustrated in the Introduction). This would give the net contribution to the observed Flynn effect from changing proportions of persons in the various sibship sizes. A reference point is needed to do these calculations. Since the Flynn effect is a change of intelligence test means from older to younger cohorts, it is most natural to use the oldest cohort group as a point of reference. Thus, the estimated means due to changing proportions of sibship sizes alone, relative to the 1938–1940 birth cohort group was calculated by keeping the mean intelligence test score within each sibship size constant on the 1938–1940 levels in all the subsequent cohort groups, and letting the proportions of the different sibship sizes vary across cohorts according to their observed values. In the first cohort group (1938–1940) the observed and the estimated means are identical. The estimated population mean in the second cohort (and similarly in each of the subsequent cohorts) was calculated as: $M_2=M_{11}*p_{21}+M_{12}*p_{22}+M_{13}*p_{23}+M_{14}*p_{24}+M_{15}*p_{25}+M_{16}*p_{26}$, where the $p$'s are the observed proportions in the second cohort, and the $M$'s are the sibship means in the first cohort (1938–1940).

## 3. Results

Fig. 1 displays the observed mean scores by sibship size in13 different birth cohort groups.

Fig. 1 shows an increase in mean General Ability within all sibship sizes, and stronger in the comparatively large sibships. In sibship sizes from 1 to 3 the increase from the 1938–1940 to the 1983–1985 birth cohorts was about 5–6 IQ points. The corresponding increases in the means for 4 and 5-sibships were approximately 7–8 IQ points. In the 6+ sibships the increase was about 10–11 IQ points.

The total population means are influenced by changing family sizes across cohorts. The relative proportions (in % units) of the various sibship sizes in each of the 13 cohort groups are shown in Fig. 2.

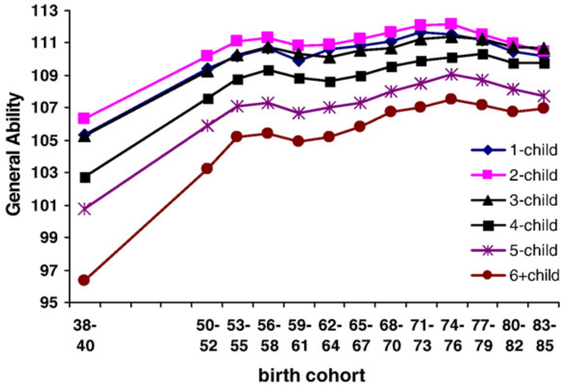

Fig. 1. Mean General Ability by sibship size and birth cohort group.

Clearly, there has been a change in the relative frequencies of the 6 different sibship sizes over birth cohorts. The general tendency is increasing proportions of small sibships at the expense of comparatively large sibships. The most dramatic feature is the drop in the relative proportion of 6+ sibships. In the 1938–1940 cohorts, the 6+ sibship size comprises nearly 20% of all sibship sizes. In the subsequent sibship sizes the proportion was, on average, less than 5%. The largest effects of proportion changes on the estimated population means should thus be expected to appear from the 1938–1940 birth cohort group to the 1950–1952 cohort group.

Fig. 3 shows the observed population means, and the estimated population means in 13 birth cohort groups, where the estimated means in each of the cohorts were calculated from the sibship means in the 1938–1940 cohort group, and the observed proportions of sibship sizes in each of the subsequent cohort groups.

As expected the increase in the estimated population means across cohorts was largest from the 1938–1940 cohort to the 1950–1952 cohort. The estimated change of intelligence in this period was around 1.5 IQ points. The change in observed means was around 6 IQ points. The estimated changes across the subsequent 12 cohorts were comparatively small (around 0.5 IQ points).

Considering the subtests, the sibship size effect was largest on the Word Similarities scores, followed by the Figures and the Arithmetic scores. In the 1938–1940 cohorts, the singletons scored on average 10.9 IQ points better than subjects from the 6+ sibship group on the Word Similarities test. The corresponding differences on the Figures and Arithmetic test were respectively 8.1 and 6.6 IQ points. In the most recent birth cohort group (1974–1985) singletons outscored the average person in 6+ sibships by 5–6 IQ points on the Word Similarities test, 1.8–2.1 on the Figures test, and 1.5–1.6 on the Arithmetic test. The changes in the sibship means in the 1974–1985 birth cohorts were comparatively small, and have been collapsed into a single group.

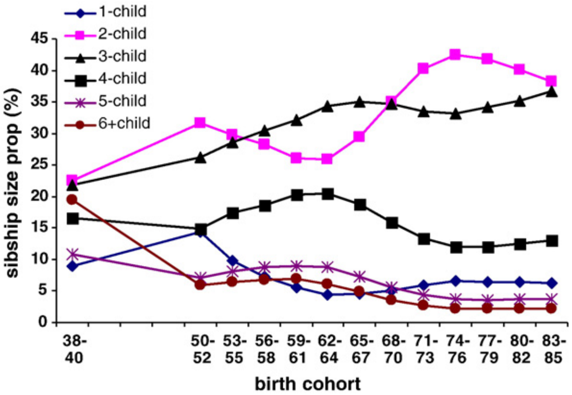

Fig. 2. Proportions (%) of persons by sibship size and birth cohort group.

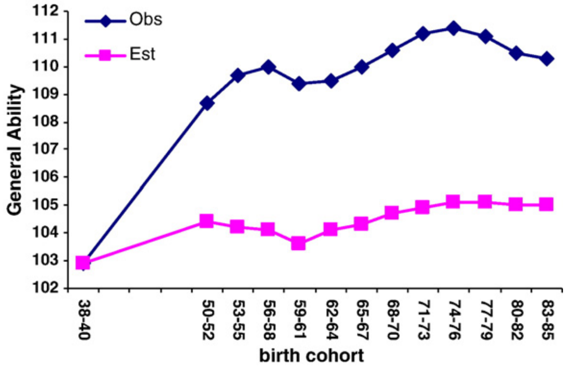

Fig. 3. Observed (obs) and estimated (est) General Ability means from changing sibship size distribution alone by birth cohort group.

Different sibship size effects on the subtests caused the estimated means (calculated in the same way as for General Ability) to be somewhat different. Table 2 displays the observed and the estimated population means for each of the three subtests.

Table 2 shows that the relation between observed and estimated means varied between the 3 subtests. Clearly, the observed means on the Figures test increased considerably more than the estimated means in the birth cohorts from 1938–1940 to 1974–1985 (around 15 and 2 IQ points, respectively). With regard to the scores on the Word Similarities test, around 35% of the observed increase from the 1938–1940 to the 1974–1985 cohorts was accounted for by the changing proportions of sibship size alone during that period. There were no appreciable changes of the observed mean scores on the Arithmetic subtest, so there is not much to explain.

Education may be a causal factor for both intelligence and sibship size, and may influence the correlation between them. In the present study, educational level (available for the 1950–1985 birth cohorts) was coded on a three-point scale according to the conventions used

Table 2
Observed (obs) mean scores and mean scores estimated (est) from changing distribution of sibship sizes alone for the Arithmetic, Figures and Word Similarities subtests by birth cohort group

| Cohort group | Arithmetic | | Figures | | Word Similarities | |
|---|---|---|---|---|---|---|
| | Obs | Est | Obs | Est | Obs | Est |
| 1938–1940 | 103.3 | 103.3 | 101.3 | 101.3 | 103.5 | 103.5 |
| 1974–1985 | 103.5 | 104.9 | 116.4 | 103.2 | 109.3 | 105.6 |
| Diff | 0.2 | 1.6 | 15.1 | 1.9 | 5.8 | 2.1 |

by the governmentally held Statistics Norway: Low educational level (compulsory school only), medium educational level (some education beyond compulsory school, but no university education) and high educational level (at least some education at the university level). The correlation between the educational level and intelligence was 0.47. Members of comparatively small sibships tended to have more education than members of larger sibships. The overall correlation between sibship size and fathers' educational level was −0.13 ($N$=940,948), decreasing from −0.16 in the oldest cohorts to around −0.05 in the more recent ones. It might be argued that crude categorization of the educational levels may have led to some underestimation of these correlations (Hunter & Schmidt, 1990). As an indirect test of this possibility, we calculated the correlation between the educational level and intelligence test scores categorized into three approximately equally large groups. The correlation between the categorized IQ scores and educational level was only slightly lower than the correlation between the continuous IQ scores and educational level ($r$=0.44). Some of the conscripts were fathers as well. We accessed around 100,000 fathers with intelligence test scores. The sons were mostly born after the early 1970's. The correlation between the intelligence of fathers and sons was 0.40, and the correlation between the intelligence of the father and the number of offspring was 0.02.

## 4. Discussion

The tests used to assess the intellectual ability among Norwegian conscripts are representative of subtests regularly included in standard intelligence tests. Thus, the Arithmetic and Word Similarities subtests were similar to the Arithmetic and Vocabulary subtests in WAIS, and the Figures test was explicitly constructed to be similar to the Raven Progressive Matrices test. In Cattell's (1987) system the first two tests measure crystallized intelligence, whereas the last one measures fluid intelligence. According to Carroll's (1993) three-stratum taxonomy, they comprise part of the second stratum, immediately below the third-stratum *g*-factor. The General Ability score, obtained by combining the scores on the three tests, correlates highly with WAIS IQ scores. The General Ability scores reported in the present paper are thus quite comparable to IQ scores obtained on standard intelligence tests.

General Ability scores have been attained for almost 1 million conscripts comprising a large part of the Norwegian (male) population over almost half a century. Subtest data were ascertained for a subset of more than 212,000 persons. A great majority of the persons were tested between their 18th and 20th birthdays, ascertaining that the secular increases in mean intelligence test scores observed in the present data set do not comprise age effects. Data on sibship size were ascertained for over 99% of the persons with intelligence test scores.

Availability of family data enables a decomposition of the population means of intelligence test scores into effects of sibship size means and the effect of the proportions of persons comprising the various sibship sizes across cohorts. Cohort changes in sibship size intelligence means and changing proportions of persons comprising the different sibship sizes can then be studied separately (Figs. 1 and 2). Estimated cohort means can be calculated by keeping one of them constant at the level of a given cohort, while calculating the estimated mean intelligence test scores by letting the other vary according to the observed data. The possibility that fertility changes is a causal factor in the secular changes of intelligence test scores has occasionally been discussed in the literature (e.g. Zajonc & Mullally, 1997). Thus, we have chosen to calculate the estimated cohort means by simulating a situation in which the mean intelligence test scores within each sibship are at a constant level across cohorts and the proportions of the different sibship sizes vary according to the observed values.

The overall impression from the present results quite clearly indicate that the increase of the relative frequency of small families at the expense of larger ones alone accounts for a distinct, but modest part of the Flynn effect both in General Ability (Fig. 3) and in the subtests (Table 2). The implication seems to be that changing fertility patterns should be taken into account when searching for the causes of the Flynn effect. It is a matter of preference whether the mean changes due to proportion changes should be considered a cause of the Flynn effect, or whether they should be considered as aggregation artifacts. The trend towards smaller families has probably been quite universal in the industrialized countries during the last 4–5 decades, and probably has contributed to the Flynn effect in all these countries. It remains to be shown whether similar family size changes may have contributed to the Flynn effect in non-industrialized countries like Kenya (Daley et al., 2003). It might be conjectured, however, that the trend towards smaller families eventually will appear also in non-industrialized countries. To the extent that the effect of sibship size on the mean intelligence test scores within a sibship is in the same direction as that observed in the present paper (Fig. 1) and other large-scale studies in other western countries (Belmont & Marolla, 1973), a Flynn effect might be expected due to changes in family size alone. Moreover, the increase per time unit will be faster if the family size patterns change comparatively quickly. In the case of increasing prevalence of large families at the expense of smaller families, a “negative” Flynn effect is expected.

The mere existence of sibship size effects on intelligence is a sine qua non for the effects of the proportion changes demonstrated in this paper. Without the sibship size effects, no effects of fertility changes are possible. The changes of the sibship means across cohorts (Fig. 1) are clearly informative about the part of the observed population mean changes that cannot be accounted for by proportion changes. This “residual” increase of mean scores is especially pronounced in the Raven-like Figures test, but is clearly discernible in the Word Similarities scores and also the composite General Ability scores (Fig. 3, Table 2). The changes in means across cohorts tend to be somewhat larger for comparatively large sibships (Fig. 2). In addition, there is a rather strong general increasing trend in the means across cohorts that is independent of sibship sizes. This is clearly seen in Fig. 1 by considering that the changes in sibship means are close to being parallel across wide ranges of cohorts.

Thus, there are three classes of (possibly overlapping) causes involved here. The causes of the negative correlation between sibship size and intelligence found in the present study (Fig. 1) and other large-scale studies (e.g. Belmont & Marolla, 1973) are still not completely understood. It could be that low-IQ parents transmit genes less favorable for intelligence to their offspring than more clever parents. This proposal implies a negative correlation between the intellectual level of the parents and sibship size. The correlation between the educational level of the fathers and sibship size in the present material varies across cohorts. In the older cohorts it was around −.15, decreasing towards −.05 in the more recent cohorts. The correlation between the intelligence of the fathers and sibship size was practically zero. This seems to indicate that genes may play some part in the older cohorts, but not in the more recent ones. The other possibility is that the lower mean intelligence in large families is caused by environmental factors. The resource dilution theory (Blake, 1981) proposes that parents have a limited amount of resources and that many children imply that each of them is allocated less resources, entailing that larger sibships should have lower mean IQ’s relative to smaller sibships. The confluence theory (Zajonc & Mullally, 1997), constructed to explain the alleged birth order effect implies sibship size effects as well. Physiological factors, like increasing probability of maternal immune attacks upon fetal brains in utero by increasing birth order (Foster & Archer, 1979), and decreasing supply of fatty acids (important for brain development) to later-born children during pregnancy (Al, v Houwelingen & Hornstra, 1997) also entail possible sibship size effects on intelligence. Recent contributions indicate that considering pregnancy and birth from an evolutionary perspective may be informative (e.g. Haig, 1993).

The changes of mean intelligence within sibships across cohorts in different sibship sizes (Fig. 1) are presumably due to environmental factors. The causes of the differential changes of sibship size means across cohorts (Fig. 1) are at present unknown, but must be sought among factors having different impacts in sibships of different sizes. Among the possible factors might be that parents in more recent cohorts have more time for each of many children than in earlier times.

The change of sibship size intelligence score means seems to be approximately the same across a wide range of cohorts (Fig. 1), implying that the causes of these changes are independent of sibship size. To explain these changes it is necessary to search for factors that affect persons in various sibship sizes in approximately the same way. General effects of improved education, parents becoming more conscious of the importance to stimulate their children intellectually, more access to stimulating games, and increasing access to media are among the possible causes. These and other possibilities have been extensively discussed (cf. Neisser, 1998), but no general consensus exists at present.

Relative to the observed change, the effect of changing sibship size proportions is smallest on the Raven-like Figures test, which also is the test with the lowest loading on the *g*-factor in the present material. Previous analyses of Norwegian conscript data have shown that the mean scores of the three subtests increased more or less in tandem from the birth cohorts immediately before the last World War to the birth cohorts in the early 1950’s (Flynn, 1987). In the subsequent cohorts, only the means on the Raven-similar test continued to increase, albeit with decreasing change rate, until it ceased to increase in the birth cohorts around the mid to late 1970’s (Sundet et al., 2004). The early changes of the subtests and General Ability means might have been caused by a combination of a changing *g*-factor and changing proportions of the different sibship sizes. The increasing means of the Raven-similar test in the subsequent 25 or so cohorts might not be explained by proportion changes. Since the means on the other two subtests, comprising the highest *g* loadings, did not increase during this period (Sundet et al., 2004), increasing *g* may not be a main component of the secular trends on the

Raven-similar test, unless a fluid *g* is invoked (Colom, Juan-Espinosa & Garciá, 2001). Varying factor structures across cohorts (Wicherts et al., 2004) indicate that measurements artifacts may be involved to some unknown extent.

The cessation, or even reversal of the Flynn effect observed in Norway (Fig. 3) and Denmark (Teasdale & Owen, 2005) may not be explained by changing proportions of sibship sizes (Fig. 2). The possibility that ceiling effects may have suppressed the means in more recent cohorts has been discussed (Sundet et al., 2004; Teasdale & Owen, 2005). While there might be some ceiling effects in the Norwegian data, this does not seem to be the case in Denmark. Data from the Netherlands (te Nijenhuis, de Jong, Evers, & Van der Flier, 2004) indicate that immigrants from non-industrialized countries have weaker (but improving) academic achievements than ethnic Dutch. Assuming that these results generalize to Norway, it is possible that a greater proportion of developing country conscripts in the Norwegian Army could have contributed to the cessation of the Flynn effect. If this is true, the prevalence of conscripts comprising large sibship sizes should have increased in the more recent cohorts. This is not the case (Fig. 2). A probable reason is that the first generation of immigrants was mostly not liable for military service for age, language and other reasons. The second immigration generation is largely still too young to enter the service.

An obvious limitation of the present data material is that it almost exclusively comprises males, and strictly does not generalize to the female population. However, Flynn (1998) reported similar intelligence gains among male and female Israeli conscripts. The selection due to the exclusion of physically or psychologically disabled persons may have introduced a small bias in the intelligence test score means, but this selection has probably remained unchanged over the observation period, and is not likely to have affected the observed secular trends appreciably. The test battery comprises only three separate tests. A larger number of subtests might have given a more complete picture of the nature of the Flynn effect.

A possible objection to the results presented in the present paper is that the calculation of the effects of changing proportion is dependent upon the reference point. It is true that the effects might have been somewhat different using other reference points than the sibship size differences in mean intelligence in the 1938–1940. Given that the Flynn effect is a change of the mean intelligence from older to more recent cohorts, it is natural to choose the oldest cohort group as a reference point.

There are still many unsolved questions regarding both the nature and the causes of the Flynn effect. To mention a few: Rodgers (1998) raised the question of whether the Flynn effect is a cohort or a period effect (the number of deaths in a war is an example of a period effect; people are more probable to die, more or less independent of birth cohort. This is still an unresolved issue, and can not, as far as we can see, be resolved in the present data set. Also the question of measurement bias needs further investigation (Wicherts et al., 2004). Quite a lot of work remains before the causes are clarified. The present results point to the potential utility of studying the Flynn effect in subgroups of different kinds and possible changes of the relative proportions of subgroups over time.

## Acknowledgements

We are indebted to the Norwegian Armed Forces and the Norwegian Social Science Data Services for kindly making the data used in this paper available to us. Psychologist Dag Barlaug at the Psychological services, Norwegian Armed Forces made subtest data available. Psychologists Tore M. Torjussen and Olav Storsve were very helpful in ascertaining data and information about the tests and testing procedures. Dag Erik Eilertsen was very helpful in organizing the data files used in the analyses. Thomas W. Teasdale and James R. Flynn generously commented on an earlier version of the present manuscript.

## References

Al, M. D. M., v Houwelingen, A. C., & Hornstra, G. (1997). Relation between birth order and the maternal and neonatal docosahexaenoic acid status. *European Journal of Clinical Nutrition*, *51*, 548−553.

Alwin, D. F. (1991). Family of origin and cohort differences in verbal ability. *American Sociological Review*, *56*, 625−638.

Belmont, L., & Marolla, F. (1973). Birth order, family size and intelligence. *Science*, *182*, 1096−1101.

Blake, J. (1981). Family size and the quality of children. *Demography*, *18*, 421−442.

Carroll, J. B. (1993). *Human cognitive abilities.* A survey of factor-analytic studies Cambridge: University Press.

Cattell, R. B. (1987). Intelligence: Its structure, growth and action Amsterdam: North-Holland.

Colom, R., Juan-Espinosa, M., & Garciá, L. F. (2001). The secular increase in test scores is a "Jensen effect". *Personality and Individual Differences*, *30*, 553−559.

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. (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, IQ gains large. *Journal of Biosocial Science*, *30*, 541−553.

Flynn, J. R. (1999). Searching for justice. The discovery of IQ gains over time. *American Psychologist*, *54*, 5−20.

Foster, J. W., & Archer, S. J. (1979). Birth order and intelligence. An immunological interpretation. *Perceptual and Motor Skills*, *48*, 79−93.

Haig, D. (1993). Genetic conflict in human pregnancy. *Quarterly Review of Biology*, *68*, 495−532.

Howard, R. W. (1999). Preliminary evidence that the average human intelligence is really rising. *Intelligence*, *27*, 235−250.

Howard, R. W. (2001). Searching the real world for signs of rising population intelligence. *Personality and Individual Differences*, *30*, 1039−1058.

Hunter, J. E., & Schmidt, F. L. (1990). Methods of meta-anlyses: Correcting error and bias in research findings London: Sage Publications.

Kane, H., & Oakland, T. H. (2000). Secular declines in Spearmans g: Some evidence from the United States. *Journal of Genetic Psychology*, *16*, 337−345.

Lynn, R., & Cooper, C. (1993). A secular decline in Spearman's g in France. *Learning and Individual Differences*, *5*, 43−48.

Lynn, R., & Hampson, S. (1986). The rise of national intelligence: Evidence from Britain, Japan and the U.S.A. *Personality and individual differences*, *7*, 23−32.

Neisser, U. (Ed.). (1998). *The Rising Curve. Long Term Gains in IQ and Related Measures* Washington DC: American Psychological Association.

Rodgers, J. L. (1998). A critique of the Flynn effect: Massive IQ gains, methological artifacts, or both? *Intelligence*, *26*, 337−356.

Rodgers, J. L., Cleveland, H. H., Van den Oord, E., & Rowe, D. C. (2000). Resolving the debate over birth order, family size and intelligence. *American Psychologist*, *55*, 599−612.

Sundet, J. M., Barlaug, D., & 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.

Sundet, J. M., Tambs, K., Harris, J. R., Magnus, P., & Torjussen, T. M. (2005). Resolving the genetic and environmental sources of the correlation between height and intelligence: A study of nearly 2600 Norwegian male twin pairs. *Twin Research and Human Genetics*, *8*, 307−311.

te Nijenhuis, J., de Jong, M. -J., Evers, A., & Van der Flier, H. (2004). Are cognitive differences between immigrants and majority groups diminishing? *European Journal of Personality*, *18*, 405−434.

Teasdale, T. W., & Owen, D. R. (1989). Continuing secular trends in intelligence and a stable prevalence of high intelligence levels. *Intelligence*, *13*, 255−262.

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

Teasdale, T. W., & 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*, 837−843.

Wicherts, J. M., Dolan, C. V., Hessen, D. J., Oosterveld, P., Van Baal, G. C., Boomsma, D. I., & Span, M. M. (2004). Are intelligence measurement invariant over time? Investigating the Flynn effect. *Intelligence*, *32*, 509−537.

Zajonc, R. B. (1976). Family configuration and intelligence. *Science*, *192*, 227−236.

Zajonc, R. B., & Mullally, P. R. (1997). Birth order: Reconciling conflicting effects. *American Psychologist*, *52*, 685−699.