Research in Developmental Disabilities 31 (2010) 1404–1411

Contents lists available at ScienceDirect

Research in Developmental Disabilities

Exploring the Flynn effect in mentally retarded adults by using a nonverbal intelligence test for children

E.E. Nijman [a], J.G.M. Scheirs [b,*], M.J.H. Prinsen [a], C.D. Abbink [a], J.B. Blok [a]

[a] Cello, Residential Centre for the Mentally Retarded, Haaren, The Netherlands
[b] Department of Psychology, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands

ARTICLE INFO

Article history:
Received 7 May 2010
Accepted 22 June 2010

Keywords:
Flynn effect
Mental retardation
SON
Nonverbal intelligence test

ABSTRACT

Increases in the scores on IQ tests across generations have been called the Flynn effect (FE). One of the unresolved questions is whether the FE affects all subsamples of the intellectual ability distribution equally. The present study was aimed at determining the size of the FE in moderately mentally retarded individuals. A nonverbal intelligence test developed for children, the Snijders-Oomen Nonverbal Intelligence Test (SON), was administered to 32 retarded adults with a mental age of 3–6 years. Sixty-nine children with a biological age in the same range and with normal intelligence served as a comparison group. Both an older and a more recent version of the SON were presented to all participants in a counterbalanced order. The proportion of items answered correctly was taken as a measure of the dependent variable. It was found that a FE existed in both the group of children and in the group of retarded adults, but that the FE was largest in the latter group. The importance of not using obsolete test norms when diagnosing mental retardation was stressed, and possible causes of the Flynn effect were discussed.

1. Introduction

It was shown by Flynn (1984) that if the same group of subjects is given two IQ tests, one normed at an earlier date than the other, subjects generally score higher on the older test. Flynn interpreted this as an indication that population IQ rises across successive generations. Later Flynn (1987) showed, by comparing the performances of distinct cohorts of people on the same test, that true increases in IQ from one generation to the next are present in a large number of industrialized countries and from 1932 onwards. A lot of research on the so-called Flynn effect has been published since, some of it extending the start of the effect backwards in time (Flynn, 1998) or expanding it to developing countries (Daley, Whaley, Sigman, Espinosa, & Neumann, 2003). Other research has even suggested that the increase in IQ has come to a standstill or a reversal in recent years (Sundet, Barlaug, & Torjussen, 2004; Teasdale & Owen, 2008). Despite this, fundamental questions remain unanswered. These questions relate to both the meaning and the cause of the effect, as was put forward in a critical paper by Rodgers (1999). To gain more fundamental insight Rodgers proposed, among other things, to investigate select samples of participants (from either the upper or lower tail of the IQ distribution), to see whether the gains not only show up in the IQ scores but also in the raw scores themselves, and to look at distributional aspects of the scores.

In this article we set out to investigate the Flynn effect in mentally retarded individuals. So far, contradictory results have been obtained in studies on the size of the Flynn effect in groups whose IQ is below average. Flynn himself concluded that the

* Corresponding author. Tel.: +31 13 4662167.
E-mail address: j.g.m.scheirs@uvt.nl (J.G.M. Scheirs).

0891-4222/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ridd.2010.06.018

effect was affecting all levels of cognitive functioning including mental retardation about equally (Flynn, 1985, 2009b, pp. 135–137). This was confirmed by Kanaya, Scullin, and Ceci (2003). Bolen, Aichinger, Hall, and Webster (1995) and Spitz (1983) found that the effect in mentally retarded individuals was there, but they did not compare different levels of intelligence. The latter author showed some years later (Spitz, 1989) that the effect diminished both above and below the average IQ range, and that the effect was even negative in the retarded range, this implying that the retarded persons' IQs were higher on the most recent test. Negative effects were also obtained by Simon and Clopton (1984) and by Goldman (1987) in moderately retarded subjects. Sanborn, Truscott, Phelps, and McDougal (2003) investigated learning disabled children and found the Flynn effect increasing from lower to higher IQ levels, although this trend was not significant. Lynn and Hampson (1986) discussed and reanalyzed several published studies but did not find a consistent relationship between size of the Flynn effect and level of intelligence. Teasdale and Owen (1989) and Sundet et al. (2004), on the other hand, compared different IQ groups and found that the Flynn effect was most pronounced at lower IQ levels.

The above results leave us with ambiguity. This might be due to the fact that the studies were not uniform with regard to the ages of the participants or to what was considered ''low IQ'' or ''retarded''. But the inconsistent results might also be due to the choice of the measuring instrument. Wechsler scales or other IQ tests developed for their use in the general population were applied in most of the investigations carried out so far. These instruments might do well in samples of children or adults with an average ability, but they probably do not provide a sensitive measurement of IQ in those with mental retardation. This is because low or very low IQs are outside the range of IQs for which the test was normed (e.g. Wechsler, 1991, 1997). Another reason is that, by definition, persons with intellectual disability will find either the items or the instructions of general tests, or both, difficult and obtain scores that cluster in the lower range of the scale, with low discriminating power and low validity of the test outcomes as a consequence (e.g., see Walsh et al., 2007). In addition, the capability of retarded persons to comprehend language is often specifically affected (Merrill, Lookadoo, & Rilea, 2003); they achieve language levels that are either consistent with or, more commonly, below what their mental age would predict (Fowler, 1998). When items in a list have to be remembered, persons with mental retardation, like the general population, tend to perform better when the items are presented in visual than in verbal form (Cherry, Applegate, & Reese, 2002).

In the present research, we chose to use a test that fits in with the capabilities and with the mental age rather than with the biological age of the participants. The test we used here consisted of two consecutive versions of the Snijders-Oomen Nonverbal Intelligence Test (SON), which was developed for use in children aged 2.5–7 years. We tested both a group of individuals with intellectual disability who were biologically older but functioning mentally at that age level, and a group of normally gifted children within the given age range. An advantage of the SON test for use with individuals who evinced intellectual disability, apart from the fact that its difficulty level is in keeping with the respondents' capabilities, is that performance on the test does not rely on language skills. The instructions and examples given to the subjects are both extensive and nonverbal, so that they can easily be understood. Feedback is given to the subject during the process of test taking and the test has both an individual starting point and a point of termination after several wrong responses, which keeps administration time limited and the respondents motivated. By its very nature, the test is aimed at measuring fluid intelligence rather than crystallized intelligence (Tellegen, Winkel, Wijnberg-Williams, & Laros, 1998). It is on tests of fluid intelligence that the largest Flynn effects have been found (Flynn, 1998).

As far as we know this is the first time that an intelligence test developed for children, and a nonverbal one, has been used to report on the Flynn effect in adults with intellectual disabilities.

2. Method

2.1. Participants

The group with intellectual disabilities had 32 adults, 11 women and 21 men, who were residents in a regional institution for the intellectually disabled in the south of the Netherlands. Their ages ranged between 28 and 60 years, with a mean of 49 and a standard deviation of 6.9 years. These persons were all moderately retarded (with estimated IQ scores ranging between 35 and 55), and on the basis of clinical judgement they were considered to function mentally at an age level of 3–6 years. This level of cognitive functioning was confirmed by the outcome of the most recent version of the SON test (to be described in the next section), showing that the so-called reference age of the participants, i.e. the age of children for whom a total test score as actually attained by the adults would lead to an IQ of 100 (Tellegen et al., 1998), was 4.5 years in this group on the average.

The participants in the second group were 69 pupils of an elementary school in a large town in the south of the Netherlands. The children in this group (28 girls and 41 boys) were between 4.1 and 6.9 years of age. The mean age was 5.3 years, with a standard deviation of .65 years. The mean reference age according to the most recent version of the SON test was 5.9 years for this group, which indicates that the children with intellectual disabilities were somewhat more advanced than the adults and also this group of children performed better than could be expected on the basis of their biological age. In other words, their average IQ must also have been above 100. This, however, might have been caused by a Flynn effect in itself, since the most recent version of the SON test dated from 1996 (see next section), while both children and adults were tested in the period fall 2006 to spring 2008.

2.2. Materials

Two versions of the SON test were used in this study. The older version, the SON 2.5-7, dates from 1975 (Snijders & Snijders-Oomen, 1976) and the revised version, the SON-R 2.5-7, dates from 1996 (Tellegen et al., 1998). The subtests of the SON are: sorting, mosaics, combining, memory and copying. The SON-R is mainly an expansion, redesigning and regrouping of the tasks that were already present in the SON, with the exception of the memory task, that is no longer part of the SON-R. The SON-R is composed of six subtests, which contain 15 items on the average. One half of the subtests is intended to measure reasoning, while the other half is intended to measure spatial performance. The six subtests and a global description of the tasks to be performed are (Tellegen et al., 1998):

1. *Mosaics*: Copying mosaic patterns in a frame, using coloured squares.
2. *Categories*: Sorting cards into two groups or choosing cards that have the same thing in common.
3. *Puzzles*: Puzzle pieces must be laid to resemble a given example or form a whole.
4. *Analogies*: Sorting disks on the basis of form and/or colour and/or size or solving an analogy problem based on these characteristics.
5. *Situations*: Choosing the correct, missing halves of incomplete drawings.
6. *Patterns*: Copying a simple pattern by drawing, or copying a pattern by connecting dotted lines.

From the subtest scores and the respondent's age, a standardized total score or SON IQ can be determined. In an evaluation of the test by the test committee of the Dutch Institute for Psychologists, the SON-R 2.5-7 was judged to be "good" with regard to test norms, reliability and both construct and criterion validity (Evers, Van Vliet-Mulder, & Groot, 2000). The older version, the SON 2.5-7, was judged to be good with regard to test norms, sufficient with regard to validity, but unsufficient with regard to reliability (Evers, Van Vliet-Mulder, & Ter Laak, 1992). It is unknown whether the reliability and validity measures as described for the general population are the same in the mentally retarded population.

2.3. Procedure

The research was carried out according to a two-group within-subjects design. The groups were adults with intellectual disabilities and children, respectively, while the repeated measurements part consisted of the SON and the SON-R being administered to all participants. The order of the two test versions was counterbalanced, implying that half of the participants in both groups received the SON first and the other half received the SON-R first. Presentation order thus constituted the second between-subjects factor. The time that elapsed between the two measurements varied between 4 and 40 days and was about 4 weeks on the average.

It was impossible to use IQ scores as a measure of the dependent variable (test performance), since the SON and SON-R have only been normed for children and IQs are not available for adults. We therefore decided to use the proportion of items answered correctly (P) as such a measure. For each individual, P was calculated per subtest first. We then had to combine the subtest scores into a P value indicating overall test performance for each subject. Since subtest results do not necessarily contribute to an overall test score (or an IQ score) equally, it was decided to do a multiple regression analysis (MRA) and to regress the IQ scores of the children, which were available for this group, on their subtest scores and age. The regression coefficients obtained in this manner were then used for weighting the subtest P values in calculating the overall P values, both in the group of children and in the group of adults. Results of the MRA will be presented in the next section.

By choosing P as the dependent variable, we abandoned the possibility of using even another measure of overall test performance, which was the reference age of the participants. Although using reference ages when IQs are not available might be appealing at first sight, this could easily lead to misleading interpretations. This is so because reference age is not defined and calculated in the same manner in both versions of the test. In the SON reference age is the age at which the obtained score is the median of the raw scores for all individuals at that age (Snijders & Snijders-Oomen, 1976), while in the SON-R it is the age at which an individual with the obtained score would be given an IQ score of 100 (Tellegen et al., 1998). The tables for converting raw scores to reference ages are also imprecise, especially in the SON, as these tables sometimes couple a range of test outcomes to the same reference age. When we did calculate the children's reference ages nevertheless and compared these to the IQ scores that were also available for this group, we found opposite effects and inexplicable results.

It was therefore decided to adhere to the P values. By doing so, we also kept close to the raw data as was suggested by Rodgers (1999).

3. Results

The results of the MRA for the SON and the SON-R in the group of children are given in Table 1.

We used the standardized coefficients as weights to calculate the overall P values. The proportion of variance explained (R^2) indicates that SON-R IQ was almost perfectly predicted by the subtest scores and the ages of the subjects. The standardized regression coefficients revealed that age was the most important predictor, while the contributions of the subtests were equally important. In the older test version, the SON, the subtests contributed unequally to IQ. Furthermore for

Table 1
Outcomes of multiple regression analyses using the children's subtest scores on the SON-R and the SON, along with their ages to predict IQ. The standardized coefficients and the proportions of variance explained are printed bold.

SON-R	*b*	**B**	*t* (61)	*p*
Mosaics	1.743	**.264**	10.68	.00
Categories	1.998	**.260**	14.76	.00
Puzzles	2.336	**.321**	16.50	.00
Analogies	1.589	**.309**	14.21	.00
Situations	2.102	**.300**	15.59	.00
Patterns	1.960	**.326**	12.15	.00
Age	−1.715	−**.988**	−45.11	.00
R^2 = .99				
SON	*b*	**B**	*t* (62)	*p*
Sorting	2.189	**.229**	4.30	.00
Mosaics	.665	**.157**	2.18	.03
Combining	2.527	**.202**	3.93	.00
Memory	1.884	**.490**	8.88	.00
Copying	3.872	**.614**	9.71	.00
Age	−1.542	−**1.112**	−16.58	.00
R^2 = .87				

that test, there was a less perfect fit between the IQ scores read from the tables and those predicted from the MRA. We conclude from this discrepancy that the SON, apart from it being possibly obsolete due to a Flynn effect, is not a very accurate instrument.

The averaged *P* values resulting from the weighted calculations, which for the SON-R did not differ from the unweighted values due to the similarity of the regression coefficients, are given in Table 2. It appears that both groups of participants performed better on the SON than on the SON-R (*P* values 0.75 vs. 0.69 for the children and 0.65 vs. 0.51 for the adults, respectively). To test for the differences, analyses of variance were first performed on the children and adult data separately. For both the children and the adults the main effect of the factor "Type of test" (SON vs. SON-R) was significant: $F(1,67) = 38.40$, $p = .000$, and $F(1,30) = 137.89$, $p = .000$, respectively. The interaction effect "Type of test × Test order" was also significant in both groups: $F(1,67) = 4.55$, $p = .037$ for the children, and $F(1,30) = 4.78$, $p = .037$ for the adults, respectively. The latter results point to the presence of a differential learning effect, i.e. to the fact that the difference between the two tests was larger in case the more recent SON-R was presented first, and that it was smaller in case the older version of the test, the SON, was presented first. In other words and not surprisingly, both children and adults learned most from the more sophisticated test.

When the data of the two groups were taken together in a larger analysis containing the factor "Group" (children vs. adults) in addition to the factors "Type of test" and "Test order", the main effects of "Group" and "Type of test" were significant. All second- and third-order interactions with "Type of test" proved also to be significant, the most interesting of these effects being the "Group × Type of test" interaction: $F(1,97) = 22.4$, $p = .000$. This test outcome indicated that the size of the difference between SON and SON-R scores was significantly larger in the adults than in the children. How to interpret the differences between the raw score averages of both tests and how to relate them to the Flynn effect, will be discussed in the next section.

When proportions are used as the dependent variable, it should be noted that such scores are restricted to the interval [0,1]. A change in a proportion from 0.5 to 0.6 is relatively smaller than a change from 0.8 to 0.9. To take this into account, we calculated the odds ratio's associated with the proportional changes. Odds ratio's above 1 indicate higher *P* values for the SON, and an odds ratio equal to 1 would indicate no difference between the two test versions. The mean odds ratio for the adults was 2.62, while for the children it was 1.97. Since distributions of odds ratio's are skewed to the right, we calculated the median values as well. The median odds ratio was 2.05 for the adults and 1.37 for the children. These outcomes corroborate the conclusion above, that persons with intellectual disabilities showed the larger effect.

Table 2
Mean scores (*P* values or proportions correct) of mentally retarded adults and children on the SON and SON-R, for two presentation orders. Standard deviations are within parentheses.

Presentation order	SON	SON-R	Total
Adults			
SON first	0.62 (0.19)	0.50 (0.18)	0.56 (0.19)
SON-R first	0.69 (0.19)	0.52 (0.16)	0.60 (0.20)
Total	0.65 (0.19)	0.51 (0.16)	0.58 (0.19)
Children			
SON first	0.72 (0.11)	0.68 (0.10)	0.70 (0.11)
SON-R first	0.78 (0.14)	0.69 (0.11)	0.73 (0.13)
Total	0.75 (0.13)	0.69 (0.10)	0.72 (0.12)

Fortunately, the conclusions based on the mean scores as presented in Table 2 and on the analyses of variance proved to be independent of our choice to weigh the subtest *P* values. When instead of the weighted *P* values, the unweighted *P* values would have been taken as the basis for the calculations and for the statistical tests, this would have led to no different patterns of mean *P* values and fundamentally to no other test outcomes.

3.1. Distributional characteristics

Distributional characteristics of the data are presented in Fig. 1. When looking at the distributions and the statistics given in this figure (standard errors of the statistics are within parentheses), it can be concluded that there were no substantial differences in spread. Neither did the skewness or kurtosis measures deviate significantly from the normal distribution, i.e. the ratio of these measures divided by their respective standard errors nowhere exceeded a critical *z* value of 2. Only the distribution of SON scores in the group of adults appeared relatively flat. The conclusion that differences between the older and newer test versions showed up in shifts of the mean and not in other moments of the distribution therefore seems warranted.

In addition to *P* values, IQ scores were available for the group of children. Mean IQ was 115.9 on the SON and 110 on the SON-R. The mean IQ of 110 on the most recent version of the test is in agreement with the mean reference age of the children. This 10-point difference in IQ (and the corresponding difference between reference age and biological age) cannot be explained by a Flynn effect alone, since there was only a 10-year period between the publication and application dates of the SON-R. This implies that our sample of children was really somewhat smarter than the general population.

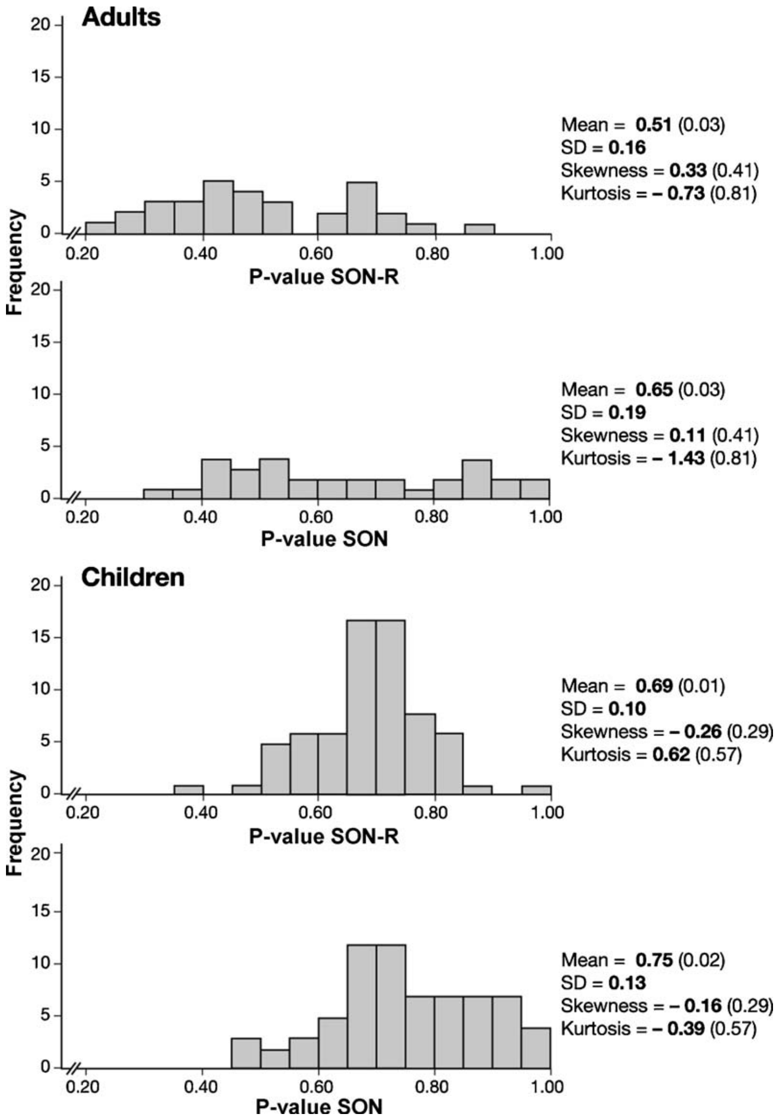

Fig. 1. Distributional characteristics of the SON-R and SON in mentally retarded adults and in children. Standard errors of the statistics (approximate values for skewness and kurtosis) are within parentheses.

4. Discussion

By using a nonverbal intelligence test developed for children in a group of children as well as in mentally retarded adults, we tried to shed more light on the occurrence and the size of the Flynn effect in the latter group.

Before asserting that this attempt succeeded, we should first answer the question how the Flynn effect is measured. This has been done in two ways. In his first publication Flynn (1984) considered the difference between an earlier and a later version of a test, both versions being given to the same subjects, as a reflection of the effect. Thereafter, Flynn (1987) applied the cross-sectional approach by comparing different cohorts of people taking the same test. Differences between test versions and differences between cohorts of people were both expressed in terms of IQ points. We did not have IQ scores here. Our data showed a diminished performance on the more recent version of the test in terms of raw scores. Such a decrease in absolute performance indicates that the more recent test was also the more difficult one. "If a test must become increasingly more difficult to reflect a normatively average score, then the latent ability that is being measured (*i.e. intelligence; authors' addition*) must be increasing as well", Rodgers (1999) stated. Since intelligence is what the Flynn effect is about, it may be concluded that a significant Flynn effect was revealed by our data in both the adults and the children. Because the increase in difficulty was larger in the adults than in the children, it may also be concluded that the Flynn effect was largest in the former group.

Rodgers (1999) pointed out that a difference in mean scores does not necessarily lead to the conclusion that IQ is changing. He showed that when a sample is measured twice, changes in variance of the population from which the sample is taken can lead to differences in mean scores also. Rodgers' arguments pertained to normed (IQ) scores, however. We did not have normed scores. We did not have changing variances in our samples of adults and children either. So, a changing variance or other changes in the distribution's shape cannot be held responsible for the results that we obtained. This leaves open the possibility that in the overall intelligence distribution, with data from several subgroups combined, a change in shape does occur. In the light of the differential mean shifts across subgroups that showed up from the present research, this is even a plausible event.

Our data showed a Flynn effect of about six IQ points over a time period of 22 years in normal 4–6-year-old children. This difference is comparable to the size of the effect in other Dutch data sets obtained from children of about the same age (Wicherts et al., 2004; Resing & Tunteler, 2007), and it fits in nicely with what has been proposed as the general size of the effect for two tests normed two decades apart (Flynn, 1998). We might have expected an even larger difference here, due to the SON tests consisting of nonverbal tasks exclusively and measuring fluid intelligence.

The present finding of a larger Flynn effect in people with a lower intelligence corroborates the results of earlier research by Lynn and Hampson (1986), Teasdale and Owen (1989), and Sundet et al. (2004). Colom, Lluis-Font, and Andrés-Pueyo (2005) also found raw score gains gradually decreasing from the bottom to the top of the distribution, but ceiling effects seem to have plagued their data. In the present research we have confirmed such a larger size of the effect specifically for moderately retarded individuals. We were able to show this on the basis of what we thought was an appropriate intelligence test. There is a dilemma here, since one can wonder if intelligence tests for children like the SON are also suited for measuring the cognitive abilities of mentally retarded adults. Admittedly, these tests have only face validity when applied to the latter group. Psychometrically sound intelligence tests designed for and normed on mentally retarded individuals do not exist, however, and tests currently in use for the general population like the Wechsler scales might have serious drawbacks when applied to the lower end of the intelligence scale (see Section 1). So, it seems that there are no better alternatives available at the moment.

What do the present results mean with regard to intellectual disability and with regard to possible explanations of the Flynn effect? First, it is clear that by using obsolete tests and test norms that have become too lenient due to the Flynn effect (which is an effect occurring automatically when tests remain in use for several years), a number of people that are actually mentally retarded will not be characterized as such. These people run the risk of not getting the special care and education that they need. This is in no way a new conclusion, but the fact that mentally retarded individuals now appear to be affected by the Flynn effect predominantly suggests that the problem might be even worse than figured out by others (e.g. Flynn, 1985; Kanaya et al., 2003).

Second, the question arises whether the present results bring us closer towards an explanation of the Flynn effect. Overviews of the alternative explanations that have been presented in the literature were, among others, given by Flynn (1998) and by Lynn (2009). The eight principal explanations according to Lynn are (1) improvement in education, (2) increased test sophistication, (3) the greater cognitive stimulation of more recent environments, (4) better child rearing practices, (5) changed attitudes on the part of the tested individuals, (6) improvement in the mean IQ of the environment in which the individual lives, (7) improvements in nutrition, and (8) the genetic factor of heterosis. Closely linked to explanations (1), (3) and (6), Flynn (2009b) added his own explanation, saying that living in an industrialized society gradually forces its inhabitants to change their way of thinking, to "put on scientific spectacles" as he called it, and to move from a concrete to an abstract way of problem solving (with increasing scores for fluid intelligence in particular as a consequence).

Lynn (2009) showed that during the second half of the 20th century there have been rates of increase of IQs of school children, of IQs of preschool children, and even of Developmental Quotients of infants aged 6–24 months similar to the rise in adult IQs. He argued that a common factor must have been the cause of all these increases over time, and of the increases in several physical characteristics that have been observed in parallel, such as infant birth weight and adult height. He identified this common factor as pre- and post-natal nutrition. Flynn (2009a) has criticized Lynn's explanation, mainly because there is not always a close correspondence between height and IQ, neither over time nor in specific groups of respondents. Yet, we feel that Lynn has a strong point. Since the Flynn effect has been shown to operate in very young children, the responsible factor must have been one that exerts its influence prenatally or very early in life. Several of the cultural influences mentioned above may then be ruled out as causes, simply because in the short history of an infant or a young child these influences do not have enough time to work.

What remains alongside the nutrition hypothesis is another early influence, which is heterosis. Heterosis is a population-wide change from a certain amount of inbreeding towards a pattern of random mating, thereby increasing the ratio of heterozygous to homozygous genotypes in the population and promoting dominant alleles only to be expressed in the phenotype. According to Mingroni (2004, 2007), any secular trend in a genetically influenced trait, IQ being one of these, could be explained by such a process.

On the basis of the present results a choice for either of the two causal mechanisms cannot be made. There is also nothing against the assumption that the two, or even other mechanisms have been working at the same time. If our results based on the SON tests are valid, however, an explanation of the Flynn effect can only be considered adequate if it accounts for the larger size of the effect in mentally retarded individuals. While it is hard to imagine that in the past the parents of persons with mental retardation moved towards a pattern of random mating more than did the general population, it is not inconceivable that the increased care for the mentally retarded led to a disproportionate improvement of their health and living conditions and possibly of their feeding conditions also (see Mans (1998, pp. 164–200, 263–289) for an overview of the fundamental changes in the care for people with mental retardation in the Netherlands from the beginning of the 20th century onwards). If future research could provide us with additional information, for instance if we knew whether individuals with intellectual disability evinced a larger increase in performance on intelligence tests accompanied by a larger increase in birth weight or adult height, then this would surely help us make decisions (e.g. in favor of a nutrition hypothesis).

It has been suggested in the literature that the Flynn effect has come to an end in recent years. One of the countries in which the rise in intelligence test scores was seen from the 1950s, then subsided and finally came to a stop in the mid-1990s, is Norway (Sundet et al., 2004). Norway is a country with a short tradition of immigration, which did not start there until 1940. Recently, the mean IQ of immigrants to Norway was found to be substantially lower than the mean national IQ (Vinogradov & Kolvereid, 2010). Selective migration might have been a factor contributing to the disappearing of the Flynn effect (see also Te Nijenhuis, De Jong, Evers, & Van der Flier, 2004). The lesson to learn is that the waxing and waning of the Flynn effect is not necessarily caused by the same factors, and also that these factors might differ from one country to another. This makes understanding the Flynn effect even more complicated. Since mechanisms such as nutrition and migration might also affect different IQ levels differentially, it seems profitable to give continued attention to the study of the effect in select samples (from either the right or left tail of the intelligence scale). Test norms may be obsolete after a decade, but Rodgers' (1999) research proposals surely are not.

Acknowledgements

We thank Prof. Wilma Resing for her valuable advice concerning data analysis in an early stage of manuscript preparation. Yvonne Janssens prepared the illustration.

References

Bolen, L. M., Aichinger, K. S., Hall, C. W., & Webster, R. E. (1995). A comparison of the performance of cognitively disabled children on the WISC-R and WISC-III. *Journal of Clinical Psychology, 51*, 89–94.

Cherry, K. E., Applegate, H., & Reese, C. M. (2002). Do adults with mental retardation show pictorial superiority effects in recall and recognition? *Research in Developmental Disabilities, 23*, 135–147.

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*, 83–91.

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.

Evers, A., Van Vliet-Mulder, J. C., & Groot, C. J. (2000). *Documentatie van tests en testresearch in Nederland*. Assen: Van Gorcum & Comp.

Evers, A., Van Vliet-Mulder, J. C., & Ter Laak, J. (1992). *Documentatie van tests en testresearch in Nederland*. Assen/Maastricht: Van Gorcum.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95*, 29–51.

Flynn, J. R. (1985). Wechsler intelligence tests: Do we really have a criterion of mental retardation? *American Journal of Mental Deficiency, 90*, 236–244.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*, 171–191.

Flynn, J. R. (1998). IQ gains over time: Toward finding the causes. In U. Neisser (Ed.), *The rising curve. Long-term gains in IQ and related measures* (pp. 25–66). Washington, DC: American Psychological Association.

Flynn, J. R. (2009a). Requiem for nutrition as the cause of IQ gains: Raven's gains in Britain 1938–2008. *Economics and Human Biology, 7*, 18–27.

Flynn, J. R. (2009b). *What is intelligence?* (Expanded paperback edition). Cambridge: Cambridge University Press.

Fowler, A. E. (1998). Language in mental retardation: Associations with and dissociations from general cognition. In J. A. Burack, R. M. Hodapp, & E. Zigler (Eds.), *Handbook of mental retardation and development* (pp. 290–333). Cambridge: Cambridge University Press.

Goldman, J. J. (1987). Differential WAIS/WAIS-R IQ discrepancies among institutionalized mentally retarded persons. *American Journal of Mental Deficiency, 91*, 633–635.

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. *American Psychologist, 58*, 778–790.

Lynn, R. (2009). What has caused the Flynn effect? Secular increases in the development quotients of infants. *Intelligence, 37*, 16–24.

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.

Mans, I. (1998). *Zin der zotheid. Vijf eeuwen cultuurgeschiedenis van zotten, onnozelen en zwakzinnigen*. Amsterdam: Uitgeverij Bert Bakker.

Merrill, E. C., Lookadoo, R., & Rilea, S. (2003). Memory, language comprehension, and mental retardation. *International Review of Research in Mental Retardation, 27*, 151–189.

Mingroni, M. A. (2004). The secular rise in IQ: Giving heterosis a closer look. *Intelligence, 32*, 65–83.

Mingroni, M. A. (2007). Resolving the IQ paradox: Heterosis as a cause of the Flynn effect and other trends. *Psychological Review, 114*, 806–829.

Resing, W. C. M., & Tunteler, E. (2007). Children becoming more intelligent: Can the Flynn effect be generalized to other child intelligence tests? *International Journal of Testing, 7*, 191–208.

Rodgers, J. L. (1999). A critique of the Flynn effect: Massive IQ gains, methodological artifacts, or both? *Intelligence, 26*, 337–356.

Sanborn, K. J., Truscott, S. D., Phelps, L., & McDougal, J. L. (2003). Does the Flynn effect differ by IQ level in samples of students classified as learning disabled? *Journal of Psychoeducational Assessment, 21*, 145–159.

Simon, C. L., & Clopton, J. R. (1984). Comparison of WAIS and WAIS-R scores of mildly and moderately mentally retarded adults. *American Journal of Mental Deficiency, 89*, 301–303.

Snijders, J., & Snijders-Oomen, Th. N. (1976). *Snijders-Oomen Non-verbal Intelligence Scale, SON 2.5-7*. Groningen: Tjeenk Willink BV.

Spitz, H. H. (1983). Intratest and intertest reliability and stability of the WISC, WISC-R, and WAIS full scale IQs in a mentally retarded population. *The Journal of Special Education, 17*, 69–80.

Spitz, H. H. (1989). Variations in Wechsler interscale IQ disparities at different levels of IQ. *Intelligence, 13*, 157–167.

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., & 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. (2008). Secular declines in cognitive test scores: A reversal of the Flynn effect. *Intelligence, 36*, 121–126.

Tellegen, P. J., Winkel, M., Wijnberg-Williams, B. J., & Laros, J. A. (1998). *Snijders-Oomen Nonverbal Intelligence Test. SON-R 2.5-7. Manual and research report*. Retrieved from http://www.testresearch.nl/sonr/sonr257manual.pdf.

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

Vinogradov, E., & Kolvereid, L. (2010). Home country national intelligence and self-employment rates among immigrants in Norway. *Intelligence, 38*, 151–159.

Walsh, D. M., Finwall, J., Touchette, P. E., McGregor, M. R., Fernandez, G. E., Lott, I. T., et al. (2007). Rapid assessment of severe cognitive impairments in individuals with developmental disabilities. *Journal of Intellectual Disability Research, 51*, 91–100.

Wechsler, D. (1991). *Wechsler intelligence scale for children* (3rd ed.). San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997). *Wechsler adult intelligence scale* (3rd ed.). San Antonio, TX: Psychological Corporation.

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