

Children Becoming More Intelligent: Can the Flynn Effect be Generalized to Other Child Intelligence Tests?

Wilma C.M. Resing
Erika Tunteler

*University of Leiden, Department of Developmental Psychology
Leiden, The Netherlands*

In this article, time effects on intelligence test scores have been investigated. In particular, we examined whether the “Flynn effect” is manifest in children from the middle and higher IQ distribution range, measured with a child intelligence test based on information processing principles—the Leiden Diagnostic Test. The test was administered to two independent groups of children—one tested recently, another tested 25 years ago. Analyses of the test scores showed that over time, children’s IQ scores increased at least 15 points for those in the higher SES group and 8 points for those in the middle SES group. In a two-factor solution, scores on the Spatial-Analytic factor increased more than scores on the Verbal factor. Inspection of the three-factor solution showed that gains on the Spatial-Analytic factor were about .9 SD, on the Memory factor .8 SD, and on the Verbal factor .6 SD. The finding that performance or fluid abilities, measuring on-the-spot problem-solving, are more susceptible to change over time than crystallized abilities, measuring the results of learning in the past, is consistent with our hypothesis. However, the finding that aspects of children’s memory are also susceptible to change over time was unexpected. It is hypothesized that the working memory system may have become more efficient over time because more information has to be absorbed at a much younger age, and coping with a stimulating environment may require a more efficient memory system.

Key words: intelligence, IQ test, Flynn effect, intellectual development

Intelligence tests—and probably other cognitive tests as well—need systematic revision from time to time. Test norms, reliabilities, and validity of tests have a time-bound character. Test norms in particular show signs of wear, for example through familiarity with test items or subtest principles, and through changes in the population. Other reasons for test revision might be the changes in scientific insights underlying test construction and test validation (e.g., Anastasi & Urbina, 1997; Cronbach, 1990; De Zeeuw, Dekker & Resing, 2004; Drenth & Sijtsma, 2006; Gregory, 2004).

In the last two decades, all these reasons for test revision have been bundled together under one big verbal umbrella by the term the “*Flynn effect*.” James Flynn introduced this term in the psychological literature (e.g., Flynn, 1984, 1987, 1994) describing a systematic and considerable rise in intelligence test scores in more than 20 countries, including the United States. The Flynn effect underlines that without revision of the test norms, people would score better and better on an intelligence test, and therefore it is necessary to revise intelligence tests at least every 10 years.

Flynn (1984) compared 73 studies in which an intelligence test had been administered twice to large samples at different points in time. Participants in these studies varied in age from 2 to 48 years. Based on this meta-analysis, Flynn concluded that the mean scores of the participants in this study showed an increase of 13.8 IQ points over a period of 46 years. This increase was visible in the test scores of participants from all age groups and within the average range of the IQ distribution. Based on a second meta-analysis on data-sets collected in 14 different countries, Flynn (1987) concluded that the observed increase in intelligence test scores had a structural, universal character. Later this rather large and influential increase on intelligence test scores over time became known as the “*Flynn effect*” (Flynn, 1999; Herrnstein & Murray, 1994), indicating that the mean level of intelligence test scores increases gradually over time.

The Flynn effect has been reported based on data-sets with both children and adolescents as participants, with test scores within the normal IQ range. Inspection of the Flynn data resulted in the conclusion that the mean increase in children’s IQs was about .50 IQ point per year (Flynn, 1984, 1987). Kanaya, Scullin, and Ceci (2003) ascertained that the Flynn effect and changes in IQ scores over time are not restricted to the average range of IQ distributions but, perhaps even more strongly, also apply to datasets based on data collected in the mentally retarded (MR) population with IQs below 70. In the Netherlands, Pesch and Ponsioen (2004) also reported Flynn effects for mildly retarded children with IQs in the lower range of the score distribution.

In most of the studies analyzed by Flynn (1984, 1987), Wechsler test versions (WISC, WISC-R, WISC-III, WAIS, WAIS-R, etc.) have been used to establish the gains in test scores over time. Kanaya et al. (2003) analyzed comparisons of WISC with WISC-R scores, and WISC-R with WISC-III scores of mentally retarded

children; Pesch and Ponsioen (2004) analyzed differences between WISC-R^{NT} and WISC-III^{NT} scores. In other studies, scores on the Raven's Progressive Matrices were compared (e.g., Teasdale & Owen, 1989; Daley, Whaley, Sigman, Espinosa, & Neumann, 2003). All the Wechsler tests have highly comparable test contents and the distinction between Verbal and Performance (V-P) subtests is common. The Wechsler tests were originally developed in 1939 and later this Verbal-Performance distinction could be interpreted in terms of the Horn-Cattell theory of fluid and crystallized intelligence (Cattell, 1943, 1963; Horn & Cattell, 1966; Kaufman, 2000). However, not all verbal subtests fit this distinction. In this theory a distinction is made between the solving of novel problems (g_f fluid intelligence) and the solving of problems with already existing knowledge acquired from schooling and culture (g_c crystallized intelligence). Within a hierarchical intelligence model, this distinction can be interpreted as a two-stratum model with g_f and g_c forming the highest, second-order level of abilities (Davidson & Downing, 2000).

Flynn (1987) came to the conclusion that the largest gains in IQ per generation were found on fluid intelligence test scores, for instance, Raven's Progressive Matrices (Raven, Court, & Raven, 1979), amounting to about 18 IQ points per generation. With a data set restricted to Wechsler tests, Flynn found larger gains on Performance scores than on Verbal scores. Flynn found gains of 11–12 IQ points per generation, and even up to 18 points, on Performance scores, which is contrary to the general expectation that scores on crystallized tests would be influenced more over time than scores on fluid intelligence tests. Language and knowledge are part of one's culture and therefore influenced by the era in which one lives and the knowledge a person already acquired in the past. However, further inspection of the test scores on a subtest level revealed that gains in IQ scores have variability in both Verbal and Performance subtests. The scores on the Verbal Wechsler subtest Similarities, for instance, showed huge increases over time. Solving Similarity problems requires both language and reasoning processes, and must, therefore, primarily be seen as "measuring on-the-spot problem-solving without the help of a previously learned method" (Flynn, 2006, p. 10), although already existing knowledge, in the form of vocabulary and other language aspects, certainly plays a role in the problem-solving process as well.

It might be questioned whether the effects reported by Flynn depend on the types of IQ tests used or the theoretical model on which they are based. In this article, our research question focused on the constancy of intelligence test scores over time, measured by a non-Wechsler intelligence test. The test used in our study, the Leidse Diagnostische Test (Leiden Diagnostic Test [LDT] (Schroots & Van Alphen de Veer, 1984), has been constructed for young children aged 4–8 years. The LDT is based on a developmental information processing model, the Mark model, in which both modalities (channels through which information is processed; for example, the auditory-speech motor and the visual-fine motor chan-

nels), and functions (information processing activities of the central nervous system) are organized hierarchically (Mark, 1962; Schroots, 1979). The factor structure of the test can be interpreted in terms of either a 2-stratum or a 3-stratum theory of intelligence (Carroll, 1993; Davidson & Downing, 2000), including a strong memory component and a general, second-order stratum.

In the underlying explorative study we examined whether the so-called Flynn effect is manifest in a group of children with IQs within the middle and higher range of the IQ score distribution, who were tested with the LDT 25 years after the construction of the test norms. Therefore, the mean LDT scores of this group of young children were compared with those of the norm group collected 25 years ago. We hypothesized that the group of children recently tested would have both higher mean IQs and higher scores on all subtests of the LDT than the children of the norm group. Based on the Flynn studies mentioned before, we also hypothesized that the increase in IQ would be between 7.5 (0.3 IQ point \times 5 years) and 12.5 (0.5 IQ-point \times 25 years) points.

METHOD

Participants

Participants in this study were 69 children (38 girls and 31 boys) from primary schools and kindergartens. The children's mean age was 6 years and 2 months (age range = 4.05–7.07 years). They came from 69 schools located all over the Netherlands. The children were tested by masters students of the Department Developmental Psychology of University of Leiden, finishing their advanced course on psychodiagnostics. The sample can be seen as representative for middle and higher socio-economic groups (SES) because the children were selected from the student-examiner's direct environment but excluded their own families and almost all student-examiners came from relatively high SES backgrounds. Moreover, the children were recruited from very different regions in the Netherlands.

Procedure

All children were tested individually, in their schools, with a Dutch child intelligence test, the LDT test developed by Schroots and Van Alphen de Veer (1984). The students who administered the test were trained in child assessment (4–12 year-old children), and had already practiced assessment with the LDT and other child intelligence tests with feedback sessions several times before administering the test. Each student-examiner assessed one child. All scoring forms were checked by two of the authors independently, and no data were missing or unclear.

Additionally, information about the educational and occupational level of both parents was collected.

Given the fact that the scores of the current sample on the LDT had to be compared with those of the original Dutch LDT-norm sample (data presented by Schroots & Van Alphen de Veer, 1984), a check was made for possible discrepancies between the two samples with respect to SES. In accordance with the classification of the SES data from the LDT-norm sample, the educational and occupational level of the father of the child were chosen to be the main characteristics for the SES classification. If there was no father, or if no data about the child's father were available, the data of the child's mother were used.

The SES classifications of the current sample displayed in Table 1 show 37 children (55%) with a high SES background (SES-I: High), consisting of children with parents with high socio-economic status, 28 children (42%) with a medium SES background (SES-II: Medium), consisting of children with parents with a middle socio-economic background, and only 2 children (3%) with a low SES (SES-III: Low) background, consisting of the children with parents with low socio-economic status.

Table 1 also shows that the current sample differs considerably from the LDT-norm sample with respect to SES. For SES-I: High, the difference in percentages of children is 40%, with more children from a high SES background in the current sample. For SES-II: Middle, Table 1 shows equal percentages of children in the two samples (42%), and for SES-III: Low the difference in percentages of children is 40%, with more children from a low SES background in the LDT-norm sample. Because these observations indicate that most of the children we tested came from relatively high SES backgrounds, it was decided to make corrections for SES in the results section of this article.

Unfortunately, in the original LDT manual (Schroots & Van Alphen de Veer, 1984) no IQ by SES data were published. To compare the data controlled for SES, we had at our disposal published detailed analyses of an experimental version of

TABLE 1
Sample characteristics: Percentages Per Socio-Economic Status Category
for the Current, LDT Norm, and LDT-E Samples

<i>SES Status</i>	<i>Percentage of Children</i>		
	<i>Current Sample^a</i> (<i>N</i> = 69)	<i>LDT-Norm Sample^b</i> (<i>N</i> = 1169)	<i>LDT-E Sample^c</i> (<i>N</i> = 411)
SES-I: high	55% (37)	15% (150)	23% (94)
SES-II: middle	42% (28)	42% (405)	27% (112)
SES-III: low	3% (2)	43% (414)	50% (205)

Note. Valid percentages (without SES = unknown) are shown.

^aSES unknown = 2. ^bSES unknown = 200. ^cSES unknown = 0.

TABLE 2
 Sample Characteristics: Percentages for Socio-Economic Status (SES)
 Category for the Current, LDT-Norm, and LDT-E Samples

SES Status	Percentage of Children		
	Current Sample ^a	LDT-Norm Sample ^b	LDT-E Sample ^c
SES-I: high	57% (37)	27% (150)	46% (94)
SES-II: middle	43% (28)	73% (405)	54% (112)

Note. Valid percentages (without SES = unknown and SES-III: low) are shown.

^aN (without SES = unknown and SES-III: low) = 65. ^bN (without SES = unknown and SES-III: low) = 555. ^cN (without SES = unknown and SES-III: low) = 206.

the LDT, the LDT-experimental (LDT-E), administered to 6-year-old children (Schroots, 1979). The SES classification of the LDT-E sample is therefore also displayed in Table 1, and from these data it can be concluded that the LDT-E sample, like the LDT-norm sample, consisted of a considerably higher percentage of children from a low (SES-III: Low) socio-economic background than the current sample. Because only two of the participants in the current study were in the low SES group (SES-III: Low), it was decided to exclude the data from these two children from the analyses, and to make comparisons among the two samples for the high and middle SES children only. The sample characteristics for middle and high SES children are displayed in Table 2. This table reveals that for high and middle SES children the percentages do not differ greatly among the current and the LDT-E sample.

Material

The LDT is an intelligence test for young children (4–8 years of age) constructed and standardized in the Netherlands. The test contains eight subtests: *Block Design*, a combination of the subtests Block Design of the WPPSI (Wechsler, 1967) and the WISC (Wechsler, 1949); *Paper Folding*, an enhanced version of the subtest Paper Folding of the Hiskey-Nebraska test for Learning Aptitude (Hiskey, 1966); *Knox Cube*, a revised version of the Revised Knox Cube Test (Arthur, 1947); *Word Span*, a newly constructed subtest based on temporal sequencing from auditory to speech motor (the child has to listen to a series of words and then repeat them in the correct order); *Picture Matching*, a newly constructed subtest based on the temporal sequencing from auditory/visual to fine motor (the child has to match a recalled series of words with pictures on a card); *Sentence Imitation*: a newly constructed subtest in which the child listens to a story and then has to reproduce the story sentences one by one immediately after hearing them; *Story Questions*, a newly con-

structured subtest in which the child has to answer questions based on the story that he or she has been told in the former subtest; and *Comprehension*, a mixed and revised version of the comprehension subtests of the WISC (Wechsler, 1949) and the WPPSI (Wechsler, 1967). The test has good psychometric qualities, with mean coefficient alphas ranging from .73 to .91 and with a coefficient alpha of .94 for the LDT deviation IQ. The test has a clear factor structure, correlates highly with other Dutch intelligence tests (e.g., Tellegen, Winkel, Wijnberg-Williams and Laros, 1998, and with early learning scales, $r = .67-.74$). The test scores provide information about both the general intelligence of the child (LDT deviation IQ) and more specific cognitive abilities (Schroots & Van Alphen de Veer, 1984).

The LDT has been constructed based on the principles of the Mark model (Mark, 1962; Mark & Mark, 1971). In this model, a number of information processing activities are postulated that enable the person to both achieve, use, and collect information and combine different parts of information in order to acquire new knowledge. According to Mark (1962) information processing goes through different modalities (channels) before learning can take place. In the input phase, information is gathered through the visual, auditory, tactual, or kinesthetic modalities; in the output phase, the speech, fine or gross motor channel are used (e.g., Schroots, 1979). According to this model there are as many channels as there are input-output combinations. Additionally, Mark discerned different functions (information processing activities) within each channel—for instance orientation and reflexes, attention, eye-hand coordination, sequencing, comprehension, reasoning. He supposed that these functions are hierarchically ordered: orientation and vigilance are supposed to be simple cognitive functions, whereas comprehension is rather more complex. The LDT has been developed from the perspective that both the visual-fine motor and the auditory-speech motor channels play an important role in cognitive information processing. From this perspective, a tripartite in channels (a visual-fine motor, an auditory-speech motor, and a combined auditory/visual-fine motor channel) led to the construction of the eight different subtests, which are described previously and shown in Figure 1. In this figure, the various subtests of the LDT are described, including their modality and function characteristics (Schroots, 1979, pp. 36–37).

Factor analysis of the LDT subtests (Schroots, Akkerman, & De Groot, 1978) showed a 2-factor and a 3-factor solution. In the 2-factor solution, the subtests Block Design, Paper Folding, and Knox Cubes contributed to the first factor, the Spatial-Analytic factor; the subtests Word Span, Picture Matching, Sentence Imitation, Story Questions, and Comprehension had high loadings on a broad Verbal factor. In the 3-factor solution, the same first factor, the Spatial-Analytic factor, was found. The other factor splits up into a Verbal Comprehension factor, including Sentence Imitation, Story Questions, and Comprehension, and a Memory factor with high loadings on the subtests Knox Cubes, Word Span, Picture Matching, and Sentence Imitation.

Subtest	Modality and function characteristics	
<i>Visual-fine motor channel</i>		
Block Design	Within a certain period of time the child must construct a figure given on paper with a restricted number of wooden blocks with several colors and designs.	<ul style="list-style-type: none"> • perceptual organization • matching analysis • synthesis • abstract conceptualization ability • visual-motor coordination
Paper Folding	A square piece of paper must be folded in exactly the same way and configuration as the tester did before. The folding starts after the tester completes the task.	<ul style="list-style-type: none"> • continuous attention • memory for spatial transformations • reasoning • visual-motor coordination
Knox Cubes	The child has to watch the tester ticking a series of blocks in sequence. After the tester completes the task, the child has to tick the same blocks in the same sequence.	<ul style="list-style-type: none"> • continuous attention • short term memory for spatial sequences • visual-motor coordination
<i>Auditory-speech motor channel</i>		
Word Span	Verbally repeating a series of words in exactly the same order as spoken by the tester.	<ul style="list-style-type: none"> • continuous attention • short term memory for words
Sentence Imitation	Verbatim reproduction of all sentences of a story read by the tester.	<ul style="list-style-type: none"> • continuous attention • short term memory for sentences
Story Questions	Answering questions about the story read by the tester in the former subtest.	<ul style="list-style-type: none"> • comprehension of questions • semantic memory • reproductive verbal ability
Comprehension	Answering questions about a variety of common subjects.	<ul style="list-style-type: none"> • social judgment • practical knowledge • logical reasoning • linguistic skills
<i>Auditory/visual-fine motor channel</i>		
Picture Matching	Matching recalled words with pictures arranged on a card.	<ul style="list-style-type: none"> • attention • memory for word sequences • matching

FIGURE 1 Modalities and function characteristics of the various subtests of the LDT.

Results

To get insight into a possible increase in intelligence test scores on the LDT over time, the mean IQ scores and standard deviations of the current sample had to be compared with those of an age-equivalent subsample of the LDT-norm sample. In Table 3 both raw and standardized means and standard deviations for the eight LDT subtests, the deviation IQ, and the LDT factors of the two research samples are presented.

TABLE 3
 Mean Scores and Standard Deviations for LDT Subtests, Deviation-IQ and LDT Factors (2- and 3-Factor Solutions)
 of the 6-Year-Old Children From the Current and LDT-Norm Samples

LDT IQ scores	Current Sample Mean Age = 6.02 Years (N = 65)			LDT-Norm Sample Mean Age = 6.00–6.05 Years (N = 150)			
	Raw Score		IQ Equivalent	Raw Score		IQ Equivalent	
	M	SD	M ^a	M	SD	M	SD
Block Design	20.40	4.69	107.80	18.93	4.70	100	15
Paper Folding	9.49	2.83	110.37	8.03	2.37	100	15
Knox Cubes	8.11	2.35	113.00	6.33	2.52	100	15
Word Span	8.29	1.60	103.65	8.05	1.89	100	15
Picture Matching	6.94	2.18	110.88	5.54	2.85	100	15
Sentence Imitation	102.29	22.28	107.12	98.62	17.66	100	15
Story Questions	16.29	5.02	105.11	14.25	5.43	100	15
Comprehension	19.23	4.21	111.71	16.24	5.64	100	15
LDT deviation-IQ			112.95			100	15
Spatial-Analytic (2f)			113.43			100	15
Verbal (2f)			109.82			100	15
Spatial-Analytic (3f)			113.43			100	15
Verbal Comprehension (3f)			111.74			100	15
Memory (3f)			108.80			100	15

^aAll means of the current sample are significantly higher ($p < .05$; one-tailed) than the corresponding means of the LDT-norm sample.

The mean raw scores and their standard deviations for the LDT-norm sample presented in Table 3 were extracted from the LDT manual (Schroots & Van Alphen de Veer, 1984, p. 46). As we are dealing with a norm sample, the mean standardized IQ and factor scores all equal the mean norm score of 100, with a standard deviation of 15. All subtests have mean standardized IQ equivalence scores ($M = 100$; $SD = 15$). One-sample t -tests were used to test possible differences in mean test, subtest, and factor scores among the LDT-norm sample and the current sample. The mean test scores of the current sample were compared with the norm scores of 100. All LDT scores of the current sample (IQ, factor and subtest scores) are significantly higher than their counterparts of the LDT-norm sample ($p < .05$). A first preliminary conclusion, based on the scores presented in Table 3, might be that the LDT deviation IQ, as expected, has increased almost 13 points over a period of 25 years.

In the conclusion above, however, sample differences in SES representativeness were not taken into account. Because no IQ scores by SES level were reported in the original LDT manual (Schroots & Van Alphen de Veer, 1984), we compared, for each SES category separately, the mean intelligence test scores and standard deviations of the current sample with those of an experimental LDT (LDT-E) sample published by Schroots (1979). For both SES-I: High, and SES-II: Middle, the raw subtest, factor, and intelligence test scores have been converted into norm scores by using the norm tables presented in the LDT manual (Schroots & Van Alphen de Veer, 1984). These scores have then been converted into IQ equivalent scores. Because the experimental version of the LDT test is not completely identical to the normed version of the LDT (Schroots & Van Alphen de Veer, 1984) used to test the children of the current sample, some adaptations in the test scores had to be made. The best estimation for the deviation IQ of the LDT-E sample is based on the summed score of the IQ equivalent subtest scores of Paper Folding, Knox Cubes, Picture Matching, Story Questions, and Comprehension multiplied by $8/5$. In Table 4 the means and standard deviations of the LDT deviation IQs are presented for the two samples by SES level. As can be seen in this table, children from the LDT-E sample with a higher SES had a mean LDT deviation IQ of 102, and children with a middle SES had a mean LDT deviation IQ of 99: a difference of three IQ points. Table 4 also shows us a clear picture by which the IQ scores of the children from the current and the LDT-E sample can be compared. Over a period of 25 years, IQ scores of children, based on a non-Wechsler test, have increased at least 15 points for children with a higher SES and eight points for children with a middle SES.

To find out whether or not LDT scores on a subtest level all show equal increases over time, z -scores were computed and are presented in Table 5. Inspection of the data in this table reveals large differences in IQ equivalent scores for the various subtests. The IQ equivalents for the subtests Word Span and Story Questions have increased relatively less, with gains of about .2–.3 of a standard deviation over a period of 25 years. The IQ equivalents for Picture Matching and Compre-

TABLE 4
Means and Standard Deviations for the LDT-Deviation IQ's
for the Two Samples by Socio-Economic Status (SES)

SES	LDT Deviation IQ			
	Current Sample Mean Age = 6.02 Years (N = 65)		Norm LDT-E Sample Mean Age = 6.00 Years (N = 206)	
	M	SD	M ^a	SD
SES-I: High	117.32	12.18	102	—
SES-II: Middle	107.18	13.72	99	—

^aBased on the subtest scores for Paper Folding, Knox Cubes, Picture Matching, Story Questions and Comprehension.

hension, on the contrary, have increased about three-quarters of a standard deviation, and the IQ equivalent for the subtest Knox Cubes has even increased nearly .9 of a standard deviation. A considerable number of the standard deviations of the mean subtest scores of the current sample are lower than 15 (see Table 3). This lower-than-expected variance might be an indication of one or more ceiling effects, but finer inspection of the scores did not reveal any indication of such effects. The lower variances probably are the result of the sample skewness.

To find out whether LDT scores on a factorial level also showed differential increases over time, z-scores, computing differences between factor scores of the current and the norm sample, were calculated and are presented in Table 6. This table presents z-scores for both the 2- and 3-factor solutions of the LDT. Inspection of the differences shows that scores on the Spatial-Analytic factor increase more than scores on the Verbal factor. When we look at the 3-factor solution we see that

TABLE 5
z-Scores Indicating the Differences in IQ
Equivalent Scores Per Subtest for the Two
Samples (Current and Norm)

LDT-Subtests	z
Block Design	0.52
Paper Folding	0.69
Knox Cubes	0.86
Word Span	0.24
Picture Matching	0.73
Sentence Imitation	0.47
Story Questions	0.34
Comprehension	0.78

TABLE 6
z-Scores, Indicating the Differences in Factor
Scores for the Two Samples (Current and Norm)

<i>LDT-Factors</i>	<i>z</i>
2-Factor Solution	
Spatial-Analytic	0.90
Verbal	0.65
3-Factor Solution	
Spatial-Analytic	0.90
Verbal Comprehension	0.59
Memory	0.78

the scores on the Spatial-Analytic factor show the largest gain (.90), followed by the scores on the Memory factor (.78). The scores on the Verbal Comprehension factor show, relatively, the least progress, but even this gain is more than half a standard deviation, e.g. nine IQ-equivalent points.

DISCUSSION

The 4–7 year-old children tested recently with the LDT appear to have higher deviation IQs than the age-equivalent children from the LDT-norm sample tested more than 25 years ago. After correction for SES, children's IQ scores increased, over a period of 25 years, by at least 15 points for children with a high socio-economic status and eight points for children with a middle socio-economic status. These results confirm our hypothesis that, in accordance with the results of multiple studies on the Flynn effect, the children tested recently would perform better on IQ tests than children of the norm group tested 25 years earlier. Flynn (1984) reported an overall increase in IQ scores of 0.3 IQ points yearly; for children even 0.48 to 0.54 IQ points (Flynn, 1984, 1987). The 12-point increase for a group of children with relatively high socio-economic status, found over a period of 25 years, matches the effect size reported by Flynn, although a different intelligence test—the LDT—based on information-processing theory was administered in our research. With this finding our main research question concerning the generalizability of the Flynn effect to a different type of child intelligence tests can be confirmed. Inspection of the nonmatched data with the data matched for SES showed that the rise in SES over time could only explain a small part of the general gain in IQ scores, and this finding is of interest precisely because it explains so little, and is in accordance with findings Flynn (1987) reported earlier.

Kanaya et al. (2003) already showed that the Flynn effect and changes in IQ norms are not restricted to scores within the average range of the IQ distribution,

but are also valid for scores within the lower range of the distribution. Our data show the existence of the Flynn effect for IQs from children in the middle and higher ranges of the IQ distribution, even with decreasing variability in test scores, with parents with relatively high SES levels. The intelligence test used in this study has been constructed on the basis of information processing theory (see Figure 1). According to the Mark model, several information processing channels and good working functions of the central nervous system are necessary to solve the various LDT subtests. Visual and auditory functioning have to be combined with the optimal functioning of the speech-motor and/or fine-motor systems. Comparing the LDT subtests with, for example, subtests from the Wechsler test series, it is evident that a relatively large number of LDT subtests need good functioning of the working memory (memory span and/or memory of sequences of information) and are measuring on-the-spot problem-solving. We therefore analyzed test scores at the factorial level. The increase in intelligence factor scores shows a differentiated picture. Looking at a 2-factor solution, the scores on the fluid Spatial-Analytic factor, with subtests requiring on-the-spot problem-solving (e.g., Flynn, 2006), have increased considerably more than the scores on the crystallized Verbal factor, with subtests measuring, at least partially, the result of previous learning. This finding is also in line with data reported earlier. Flynn (1987, 1999) concluded that the largest IQ increase has been found on the most fluid cognitive tests, such as certain tests within the performance factor of the Wechsler tests and verbal Wechsler tests, for instance Similarities, requiring reasoning and mental energy (Flynn, 2006), and Raven's Progressive Matrices. Looking at LDT scores from a 3-factor perspective, a finer distinction could be made within the broad Verbal factor into a Memory factor and a Verbal Comprehension factor, whereas the Spatial-Analytic factor did not change. The test scores of the children from the current sample increased about .9 of a standard deviation on the Spatial-Analytic factor, .6 of a standard deviation on the Verbal Comprehension factor, as could be expected from earlier research on the Flynn effect mentioned before, but increased, quite unexpectedly, about .75 of a standard deviation on the Memory factor, including subtests based on visual and auditory memory span, memory for spoken word sequences, and memory for fine-motor moving pattern sequences, measuring aspects of the working memory of the children. Like fluid tests, memory tests require a lot of mental energy and one expects that the test scores will not be easily influenced by previous learning at school or at home.

The finding that fluid abilities and memory are more susceptible to change than abilities which can be influenced more easily by previous learning or are part of the traditional classroom subjects is consistent with findings reported by Flynn (1998, 1999, 2006), but remains, at first glance, at odds with our logical expectations. Language and prior knowledge, as part of crystallized intelligence, are integrated in and influenced by our cultures and therefore are influenced by the era in which a person lives. Greenfield (1998) even came to the conclusion that "nonverbal IQ

tests are, in fact, more, not less culture sensitive than verbal tests” (p. 118), because verbal tests need translation and revision for each country and culture. Data within the Verbal factor, however, show that abilities underlying tests involving on-the-spot problem-solving, whether they have a verbal or nonverbal character, are much more sensitive to change than crystallized tests requiring previous learning in or outside schools (Flynn, 2006). Within the Verbal Comprehension factor, the scores on the subtest Comprehension increased most (.78), indicating that verbal reasoning—which cannot easily be influenced by former learning—plays an important role in solving the type of items in this subtest. Therefore, it is not the distinction between verbal and performance abilities that matters in the Flynn effect, but the distinction between problem-solving based on new learning and problem-solving based on previously acquired knowledge.

According to Carroll (1993), fluid intelligence and memory have been seen for a long time as abilities that are not susceptible to much influence from the outside. For instance, for a long time it was thought that spatial abilities were very resistant to training. Research by Pellegrino and Kail (1982) and Pellegrino, Alderton, and Shute (1984), among others, showed otherwise. Still, the finding that aspects of the working memory of children are susceptible to change over time remains surprising. According to most researchers in this field, central memory resources and memory capacity are limited. One explanation for our findings might be that use of the working memory system has become more efficient over the years. Children have to process much more and varied information at much younger ages than 25 years ago, and coping with a very stimulating environment might require the memory system to be more efficient. Generally spoken, there was until now no clear and uniform explanation for these memory aspects of the IQ gains over time. At least, genetic changes are to be excluded as causal factors, because relatively large IQ changes have been found even within one generation. Frequently mentioned possible explanations, such as a higher educational level, a higher socio-economic level of the population, and test sophistication probably contribute to the Flynn effect, but Flynn (1987) estimated their contribution at the most three IQ points, and our data contribute to this findings.

Neisser et al. (1996) and Neisser (1998) described three possible sources for explaining the IQ gains: First, an increasing complexity of the society we live in. Factors like urbanization influence the media, and more years of schooling will potentially result in higher IQ levels. Among others, Schooler (1998), Greenfield (1998), Williams (1998), and Ceci, Rosenblum, and Kumpf (1998) studied the influences of these factors. The reported effects are not high enough to explain the whole Flynn (1998) effect, but Schooler (1998) stated that Flynn “errs in downplaying the role of these socio-environmental changes in explaining the effect he discovered” (p. 77). After all, changes in society as mentioned above are going at least as fast as the IQ changes reported by Flynn, who came, in an article on rethinking the concept of intelligence and what effects it has, to the conclusion

that not only are each of the different aspects of our complex society important explanations for the rise in IQ scores, but the interaction between all kinds of social influences are significant, for example, the emphasis in all kinds of jobs on instantaneous, on-the-spot problem-solving, the reduction of family size, the changes in interests of children (cognitive), the importance of and attention to schooling, the increase in complex information technology, etc. (Flynn, 2006).

The second possible explanation of the IQ gains is that nutrition has improved, and better food might influence brain performance and, indirectly, mental performance. Among others, Sigman and Whaley (1998) studied the influences of better nutrition and stated that “the rise in intelligence over the last 50 years is likely to have been influenced by the improved standards of nutrition” (pp. 176–177). Lynn (1998) also stressed the importance of good nutrition for improved mental performances. According to him, fluid intelligence is much more sensitive to variations in nutrition than crystallized intelligence. Daley et al. (2003) reported Flynn effects in a developing country, Kenya (1984–1998). They found significant increases in children’s IQs over a period of 14 years and hypothesized that improvements in children’s nutrition and health combined with good family structure and literacy of the parents are the main factors explaining the Flynn effect. Flynn (2006) reviewed the literature on IQ gains and nutrition and concluded that the results of almost all studies confirm that after 1948 no evidence has been found for a correlation between (better) nutrition and IQ gains.

The third possible explanation is that IQ tests measure only part of the thinking underlying intelligence: abstract problem-solving. Flynn (1987, 1999) proposed this last hypothesis: according to him, it is the only explanation for the IQ gain phenomenon because the other variables did not appear to have enough power of explanation.

To date, most researchers in this field have not given much attention to multi-factorial explanations and to the way the various factors influence and/or enforce each other mutually, although recently some multi-factorial models have been constructed (Dickens & Flynn, 2001, 2002) and rejected (Loehlin, 2002; Rowe & Rodgers, 2002). At the same time, most of the attention of the researchers is directed at identifying possible causal factors to explain the Flynn effect. However, not much attention is given to individual or group score patterns on the various facets within the intelligence domain and to possible explanatory differences in cognitive processes lying behind the test scores. Other points of interest are (differences in) the variance patterns of intelligence test data within longitudinal research.

In the study reported here, the data were not normally spread as far as level of socio-economic status was concerned. This complicated the interpretation of the data. Future studies should avoid this problem. Furthermore, longitudinal studies in combination with cross-sectional data should give us more insight into the question of whether all people show equal gains in their IQ scores or whether some

groups of persons show more gain than others. Based on the research so far, we can conclude that there were gains in test scores over a period of a quarter century, but we cannot say anything about the linearity (or nonlinearity) of this gain. Studies with a combination of repeated longitudinal and cross-sectional measurements would be necessary. Further research should also focus on questions about how long the Flynn effect will still be found, and how large changes in the populations, for example by immigrant and refugee movements, will influence the Flynn effect in the near future.

In conclusion, it is clear that intelligence tests should be restandardized every 10 years. This is the case for both more traditional Wechsler tests and for tests like the LDT designed from an information processing perspective. Using tests with old test norms will lead to an overestimation of the intelligence of tested individuals and has therefore consequences for individual diagnostic interpretations and decisions of the diagnostician using the tests. This not only applies to individuals with test scores within the normal range of the intelligence test score distribution, but also for children with scores at both the lower and higher end of this distribution.

ACKNOWLEDGMENTS

We would like to thank Marjon Nijland for collecting part of the data of the study. We are indebted to the anonymous reviewers of this article for their valuable comments.

REFERENCES

- Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Upper Saddle River, NJ: Prentice Hall.
- Arthur, G. (1947). *A point scale of performance tests. Revised Form II. Manual for administering and scoring the tests*. New York: The Psychological Corporation.
- Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, MA: Cambridge University Press.
- Cattell, R.B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, *40*, 153–193.
- Cattell, R.B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*, 1–22.
- Ceci, S.J., Rosenblum, T.B., & Kumpf, M. (1998). The shrinking gap between high- and low-scoring groups: Current trends and possible causes. In U. Neisser (Ed.), *The rising curve* (pp. 287–302). Washington DC: American Psychological Association.
- Cronbach, L.J. (1990). *Essentials of psychological testing* (5th ed.). New York: Harper & Row.
- 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.
- Davidson, J.E., & Downing, C.L. (2000). Contemporary models of intelligence. In R.J. Sternberg (Ed.), *Handbook of intelligence* (pp. 34–52). New York: Cambridge University Press.

- DeZeeuw, J., Dekker, R., & Resing, W.C.M. (2004). *Algemene Psychodiagnostiek 1: Testmethoden* [General Psychodiagnostics 1: Assessment methods]. Leiden, The Netherlands: PITS.
- Dickens, W.T., & Flynn, J.R. (2001). Heritability estimates versus large environmental effects: The IQ Paradox resolved. *Psychological Review*, *108*, 346–369.
- Dickens, W.T., & Flynn, J.R. (2002). The IQ paradox is still resolved: Reply to Loehlin (2002) and Rowe and Rodgers (2002). *Psychological Review*, *109*, 764–771.
- Drenth, P.J.D., & Sijtsma, K. (2006). *Testtheorie* [Test theory]. Houten, The Netherlands: Bohn Stafleu van Loghum.
- 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. (1994). IQ gains over time. In R.J. Sternberg (Ed.), *Encyclopedia of human intelligence* (p. 617–623). New York: MacMillan.
- Flynn, J.R. (1998). IQ gains over time: Toward finding the causes. In U. Neisser (Ed.), *The rising curve* (pp. 25–66). Washington DC: American Psychological Association.
- Flynn, J.R. (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist*, *54*, 5–20.
- Flynn, J. R. (2006). Efeito Flynn: Repensando a inteligência e seus efeitos [The Flynn Effect: Re-thinking intelligence and what effects it]. In C.F.-M.R. Colom (Ed.), *Introdução à Psicologia das Diferenças Individuais* [Introduction to the Psychology of Individual Differences] (pp. 387–411). Porto Alegre, Brazil: ArtMed.
- Greenfield, P.M. (1998). The cultural evolution of IQ. In U. Neisser (Ed.), *The rising curve* (pp. 81–123). Washington DC: American Psychological Association.
- Gregory, R.J. (2004). *Psychological testing. History, principles and applications*. Boston: Pearson Education Group.
- Herrnstein, R.J., & Murray, C. (1994). *The Bell curve: Intelligence and class structure in American life*. New York: Free Press.
- Hiskey, M.S. (1966). *Manual for the Hiskey-Nebraska Test of Learning Aptitude (HNTLA)*. Lincoln, NB: Union College Press.
- Horn, J.L., & Cattell, R.B. (1966). Refinement of the theory of fluid and crystallized general intelligence. *Journal of Educational Psychology*, *57*, 253–270.
- 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.
- Kaufman, A.S. (2000). Tests of intelligence. In R.J. Sternberg (Ed.), *Handbook of intelligence* (pp. 445–476). New York: Cambridge University Press.
- Loehlin, J.C. (2002). Comments—The IQ paradox resolved? Still an open question. *Psychological Review*, *109*, 754–758.
- Lynn, R. (1998). In support of the nutrition theory. In U. Neisser (Ed.), *The rising curve* (pp. 207–215). Washington, DC: American Psychological Association.
- Mark, H.J. (1962). Elementary thinking and the classification of behavior. *Science*, *135*, 75–87.
- Mark, H.J., & Mark, S.A. (1971). Neuropsychological examinations in young children. In J. Hellmuth (Ed.), *Exceptional infant. Vol.2: Studies in abnormalities* (pp. 41–53). New York: Brunner.
- Neisser, U. (Ed.). (1998). *The rising curve*. Washington, DC: American Psychological Association.
- Neisser, U., Boodoo, G., Bouchard, T.J., Jr., Boykin, A.W., Brody, N., Ceci, S.J., Halpern, D.F., Loehlin, J.C., Perloff, R., Sternberg, R.J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*, 77–101.
- Pellegrino, J.W., Alderton, D.L., & Shute, V.J. (1984). Understanding spatial ability. *Educational Psychology*, *19*, 239–253.

- Pellegrino, J.W., & Kail, R., Jr. (1982). Process analyses of spatial aptitude. In R.J. Sternberg (Ed.), *Advances in the psychology of intelligence*, Vol. 1 (pp. 311–365). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Pesch, J.W.E., & Ponsioen, A.J.G.B. (2004). Flinterdunne en flagrante Flynn-effecten bij licht verstandelijk gehandicapte kinderen [Tiny and flagrant Flynn effects in mildly mentally retarded children]. *De Psycholoog*, 39, 64–68.
- Raven, J.C., Court, J.H., & Raven, J. (1979). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. London: H.K. Lewis & Co.
- Rowe, D.C., & Rodgers, J.L. (2002). Expanding variance and the case of historical changes in IQ means: A critique of Dickens and Flynn (2001). *Psychological Review*, 109, 759–763.
- Schooler, C. (1998). Environmental complexity and the Flynn effect. In U. Neisser (Ed.), *The rising curve* (pp. 67–79). Washington DC: American Psychological Association.
- Schroots, J.J.F. (1979). *Leidse Diagnostische Test: Deel 5/Cognitieve ontwikkeling, leervermogen en schoolprestaties* [Leiden Diagnostics Test: Part 5: Cognitive development, learning abilities and school achievements]. Lisse, The Netherlands: Swets & Zeitlinger.
- Schroots, J.J.F., Akkerman, A.E. & De Groot, A. (1978). *Leidse Diagnostische Test: Deel 2/ Handleiding voor het stroomdiagram; van klinisch oordeel tot psychologisch rapport* [Leiden Diagnostics Test: Part 2/Manual for the flow diagram; from clinical assessment to psychological report]. Lisse, The Netherlands: Swets & Zeitlinger.
- Schroots, J.J.F. & Van Alphen de Veer, R.J. (1984). *Leidse Diagnostische Test: Deel 1/handleiding* (3^e druk) [Leiden Diagnostics Test: Part 1/Manual (3rd ed.)]. Lisse: Swets & Zeitlinger.
- Sigman, M., & Whaley, S.E. (1998). The role of nutrition in the development of intelligence. In U. Neisser (Ed.), *The rising curve* (pp. 155–182). Washington DC: American Psychological Association.
- Teasdale, T.W., & Owen, D.R. (1989). Continued secular increases in intelligence and a stable prevalence of high intelligence level. *Intelligence*, 13, 255–262.
- Tellegen, P.J., Winkel, M., Wijnberg-Williams, B.J., & Laros, J. (1998). *Snijders-Oomen Niet-Verbale intelligentietest SON-R 2 ½-7. Handleiding en verantwoording*. [Snijders-Oomen Non-verbal Intelligence test SON-R 2 ½-7. Manual]. Göttingen, Germany: Hogrefe Verlag.
- Wechsler, D. (1949). *Manual for the Wechsler Intelligence Scale for children*. New York: The Psychological Corporation.
- Wechsler, D. (1967). *Manual for the Wechsler Preschool and Primary Scale of intelligence*. New York: The Psychological Corporation.
- Williams, W.M. (1998). Are we raising smarter children today? School- and home-related influences on IQ. In U. Neisser (Ed.), *The rising curve* (pp. 125–154). Washington DC: American Psychological Association.

Copyright of International Journal of Testing is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.