Available online at www.sciencedirect.com
ScienceDirect
Intelligence 35 (2007) 187–196
INTELLIGENCE

Identification of a Flynn Effect in the NLSY: Moving from the center to the boundaries

Joseph Lee Rodgers *, Linda Wänström

Department of Psychology, University of Oklahoma, Norman, OK 73019, USA

Received 17 January 2006; received in revised form 5 April 2006; accepted 19 June 2006
Available online 31 July 2006

Abstract

The Flynn Effect [Flynn, J.R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. Psychological Bulletin 95, 29–51.] is an increase in IQ of around .33 points per year, observed in developed (and some developing) countries during the past century. It emerges from problem solving and other non-verbal components of IQ. The cause has been argued and theories proposed. Rodgers [Rodgers, J.L. (1998). A critique of the Flynn Effect: Massive IQ gains, methodological artifacts, or both? Intelligence 26, 337–356.] noted that the search for causes has preceded specification of the nature of the effect. Our study uses a national sample of U.S. children to test for the Flynn Effect in PIAT-Math, PIAT-Reading Recognition, PIAT-Reading Comprehension, Digit Span, and PPVT. An effect of the predicted magnitude was observed for PIAT-Math when maternal IQ was controlled. This finding in a large representative sample with thousands of variables supports more careful evaluation of the Flynn Effect, in demographic, geographic, environmental, and biological domains.

Keywords: Flynn Effect; PPVT; PIAT-Math; PIAT-Reading Recognition; PIAT-Reading Comprehension; IQ changes; NLSY; NLSY-children

The basic empirical status of the Flynn Effect is well-established. In developed countries, measured IQ has been increasing at a rate of approximately .33 IQ points per year for most of the past century, a pattern consistent across both time and geography. Flynn (1984) documented the trend in U.S. IQ data and then showed its existence in 14 developed countries (Flynn, 1987). The effect has recently been shown in developing countries as well (Daley, Whaley, Sigman, Espinosa, & Neumann, 2003). The increase occurs primarily within the problem solving/spatial reasoning domain, as captured by culture free tests (e.g., the Raven's Progressive Matrices) — i.e., for fluid intelligence (Horn & Cattell, 1966). Verbal IQ – crystallized intelligence – has been stable or may even have fallen slightly during this period (for discussion, see Flynn, 1987, 2006; Jensen, 1991; Loehlin, 1996).

But aside from these basic empirical findings, a great deal is *not* known about the Flynn Effect. There is no consensual agreement as to the cause of the Flynn Effect, though many have contended to provide an ultimate explanation. Theories include nutritional improvements (Lynn, 1998; Martorell, 1998), exposure to movies and optical displays (Greenfield, 1998), the increased use of speeded tests (Brand, 1996), an ever-broadening cultural knowledge base that emerges from collective memory (Mahlberg, 1997), schooling and educational methods (Blair, Gmson, Thorne, & Baker, 2005), pre-school education in particular (Teasdale & Berliner, 1991), and

* Corresponding author.
E-mail address: jrodgers@ou.edu (J.L. Rodgers).

0160-2896/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2006.06.002

a multiplicity hypothesis suggesting a great many small causes that combine together (Jensen, 1996).

Many argue that the Flynn Effect necessarily must have environmental, and not biological/genetic origins — could the genome change so rapidly and consistently? But this belief yields a paradox, given that cognitive performance scores have a moderate-to-high and stable heritability and low shared environmental variance (e.g., Loehlin, 1989; Rodgers, Rowe, & May, 1994). Dickens and Flynn (2001) referred to this paradoxical cause as "Factor X" and proposed a dynamic gene–environment interaction model to explain the paradox; the dynamic features of their model allow small environmental changes to magnify into large influences on IQ (also see critiques by Loehlin, 2002; Rowe & Rodgers, 2002). Mingroni (2004) suggested a genetic-based hypothesis based on heterosis (also called hybrid vigor); as the ratio of heterozygous to homozygous genotypes increases in the population due to broadening mating patterns, IQ responds positively (see also Jensen, 1998). Kanaya, Ceci, and Scullin (2005) used a growth-curve modeling approach to investigate the interaction of age with IQ norming patterns in relation to the Flynn Effect, and recommended more attention to longitudinal research and accounting for age in future investigations of the Flynn Effect.

Earlier, Flynn himself was not sure that the Flynn Effect has much to do with intelligence; for example, Flynn (1996) consistently referred to "ersatz intelligence gains," suggesting that the effect is an artifact. Wicherts et al. (2004) found lack of factorial invariance in the Flynn Effect across cohorts, raising fundamental questions about the stability of norming samples and even the consistent meaning of subscales and item information across cohorts. Flynn (1998) reviewed the causal hypotheses and updated his positions in Flynn (2006). There he begins to rebuild a case for the legitimacy of IQ. After he re-states the earlier position – "we are driven to the conclusion that massive IQ gains are not intelligence gains or, indeed, any kind of significant cognitive gains" – he follows by suggesting that the "best way to show that IQ gains are real is to show how they illuminate what is going on in the real world." Then he reviews the substantive import of real IQ gains in educational, economic, and cognitive domains.

Development of causal theories, though tempting, may still be premature. Rodgers (1998), in a general critique of research on the Flynn Effect, suggested that "research addressing the *legitimacy* and *meaning* of the effect should precede research *testing for* and evaluating *causes of* the effect" (p. 338, italics in original). Rodgers proposed a number of important questions that had not been adequately addressed. Does the Flynn Effect operate within individuals, within families, within cohorts? Does the effect hold up as consistent across race, gender, and ability strata? In normed or raw cognitive ability scores? In overall scale scores, in subscale scores, or in item scores? He concluded with 10 proposals for future research to specify the nature and meaning of the effect. Indeed, as substantial research attention has been devoted to the Flynn Effect, some of these questions are being carefully evaluated, and legitimate evidence brought to bear. But almost a decade later, many of the questions posed by Rodgers have not been resolved; indeed, many have not even been addressed. Prerequisite for addressing many of those questions is the existence of appropriate and powerful data and designs.

The research presented in this article provides both a motivation and a mechanism to look at the Flynn Effect more deeply and more broadly. We begin our treatment by re-framing the Flynn Effect. Then, our empirical analysis tests for the Flynn Effect in a national database that is broad enough to provide variables and context to address many previously unanswered Flynn Effect questions. We conclude by specifying a number of critical research agendas.

1. Re-framing the Flynn Effect

The Flynn Effect has been empirically identified by a number of different research efforts (e.g., Flynn, 1984, 1987; Lynn, 1990; Lynn & Hampson, 1986; Teasdale & Owen, 1987, 1989). Flynn (2006) notes that knowledge of the effect preceded his original attention. Lynn (1982) reported a Flynn Effect in a comparison of Japanese and U.S. data slightly earlier than Flynn, prompting Rushton's (1997) suggestion to rename the phenomenon the "Lynn–Flynn Effect." As Flynn (2006) noted, the effect itself was observed much earlier (see, e.g., Smith, 1942; Tuddenham, 1948). But with few exceptions, no careful and systematic accounting has been done to define the various domains – demographic, geographic, environmental, and biological – and the boundaries within each domain in which the Flynn Effect occurs. Our proposal is to move research on the Flynn Effect to those boundaries and away from the middle. Rodgers (1998) criticized the automatic interpretation of the Flynn Effect as a change in population means; there are variance (and other moment) interpretations (see Rowe & Rodgers, 2002 and responses from Dickens & Flynn, 2002 for further discussion). Our suggestion is even broader — to more sharply define the Flynn Effect by specifying the boundary conditions, which have never been systematically treated.

Some recent work on "boundary conditions" has been done. The Flynn Effect shows up in problem solving domains, and not in verbal domains, in measures of fluid intelligence rather than crystallized intelligence. Two studies suggested that the Flynn Effect occurs predominantly in the lower tail of the IQ distribution (Colom, Lluis-Font, & Andres-Pueyo, 2005; Teasdale & Owen, 1989). A study of Norway data showed the Flynn Effect, strong through most of the 20th century, dissipating since the early 1970s (Sundet, Barlaug, & Torjussen, 2004); a similar slowing was documented in Denmark as well (Teasdale & Owen, 2000). Otherwise, in what domains do these "massive gains in IQ" (Flynn, 1984) occur? In what domains do they not occur?

2. Methods

The National Longitudinal Survey of Youth (NLSY) began as a household probability sample of 12,686 adolescents aged 14–21 at the end of 1978. Those respondents have been followed on a yearly or biyearly basis ever since. In 1986, a new survey was developed based on all children born to the NLSY females, the Children of the National Longitudinal Survey of Youth (NLSYC). These children – including all new births by each survey year – have been surveyed every other year since then. By 2000, the NLSY females were age 35 to 43 and had given birth to 11,205 children (well over 90% of the projected total children who will be born to this cohort), who ranged in age from newly born to the early 20s. For our purposes, the critical feature of the NLSYC database is that a set of cognitive ability assessments were administered to all NLSYC respondents within broad age ranges; these assessments will be used to test for a Flynn Effect in the NLSYC. The NLSYC is a sample of children born to a representative (probability) sample of mothers, and thus the NLSYC contains sampling weights that can be used to provide population estimates. Sampling weights are used in all major analyses presented in this paper; results can be generalized to the population of children born to mothers who were 14–21-year-old adolescents on December 31, 1978. Because of the size of the overall NLSYC sample, statistical analyses based even on subgroups of the respondents will have high statistical power relative to most other studies and can identify relatively small effect sizes. However, the Flynn Effect has not been a small effect size in the past. The 14-year time period included within our design predicts an approximate 4.5-point increase in the context of an IQ metric.

The assessments include the three Peabody Individual Achievement Test (PIAT) subscales. The PIAT-Math (PIAT-M) test measures quantitative reasoning, the PIAT-Reading Recognition (PIAT-RR) test measures oral reading ability, and the PIAT-Reading Comprehension (PIAT-RC) test measures the ability to derive meaning from written words. PIAT scores were collected in each survey year for children aged 5–13. Also included was the Wechsler Memory for Digit Span test, which measures short-term memory of ordered numbers; Digit Span was collected for children aged 7–11. The Peabody Picture Vocabulary Test (PPVT), in which the child hears a sentence and points to the relevant picture, was administered to children aged 3–12. Of these assessments, the one predicted to show a Flynn Effect is the PIAT-M, the assessment that maps into fluid intelligence. The other assessments, more strongly verbally based, map into the crystallized domain.

The design that we use to test for the Flynn Effect provided up to nine design replications, one for each age from 5 to 13. If the Flynn Effect exists in these data, we expect to see higher test scores for 5-year-olds from the 2000 survey, compared to 5-year-olds from the 1998 survey, compared to those from the 1996 survey, etc., back to the 1986 survey. The same patterns should occur for 6-year-olds, 7-year-olds, etc. It is important to note exactly what these types of potential effects (potential Flynn Effect gains) represent within the context of this design. As in past Flynn Effect research, these potential gains occur within cross-sectional comparisons across age cohorts; those are the types of comparisons on which past Flynn Effect patterns have been based. Past Flynn Effects do *not* appear to reflect either within-person increases over time, nor do they appear across children within families (e.g., there is not an increase in IQ with increasing birth order when siblings within the same family are compared to one another; Rodgers, Cleveland, van den Oord, & Rowe, 2000; Wichman, Rodgers, & MacCallum, 2006).

However, there is an obvious threat to internal validity for which we must account, a well-known selection bias (see discussion in Chase-Lansdale, Mott, Brooks-Gunn, & Phillips, 1991; Rodgers & Rowe, 1988; among many others). In the NLSYC, because of the survey design, mothers of children who were age 5 in 1986 are younger on average than those of children age 5 in 1988, etc. This effect occurs because the mothers' age increases by 2 years between each of the NLSYC data collection cycles. Mother's IQ typically correlates positively with age-at-first-birth (AFB). This pattern has been found in a number of different studies (e.g., Retherford & Sewell, 1989), with the size of the correlation ranging from relatively small to moderate. Neiss, Rowe, and Rodgers (2002) reported several correlations between IQ and age-at-first-birth for different genetic categories in the NLSY data from the mid-1990s; the median correlation was $r_{IQ,AFB}=.27$. In the current study, we correlated maternal IQ and age-at-first-birth using the 2002 NLSY data for 3192 mothers and found $r_{IQ,AFB}=.43$. Thus, if we find an increase in child cognitive measures across years (what we would expect to interpret as evidence of the Flynn Effect), we actually could be observing the link between mother's IQ and the child's assessment scores. Fortunately, measures are available within the original NLSY files to account for this bias within our statistical models. Maternal IQ was measured in the NLSY as the Armed Forces Qualification Test (AFQT) scores (the IQ scales from the Armed Services Vocational Aptitude Battery—ASVAB). Thus, we can control for maternal IQ to adjust for this obvious and important source of selection bias. The AFQT scores were obtained in 1980 from all of the original NLSY respondents, when they were 15–23 years of age. All were referred to the same norming process, which assumes adult responses.

There are two interesting problems in using maternal IQ to control for this type of selection. First, the maternal IQ scores may themselves be subject to a Flynn Effect. We will investigate this concern empirically using the ages of the NLSY mothers in 1979 as an indicator that could potentially show a Flynn Effect. This would be evidenced as a negative correlation between mother's IQ and maternal age in 1979. (We hasten to note that the NLSY79 is not a very strong design for evaluating a Flynn Effect; because it is based on a single score from a single cross-section, there would be other validity threats if this were a substantive effort to estimate a Flynn Effect, rather than the investigation of a possible confound.) Second, because maternal IQ is correlated with child IQ, we may remove some of the useful variance in using it as a covariate. We will investigate this concern empirically also.

The other issue that we will investigate is the relationship between normed IQ scores and raw IQ scores. Because of the design of our study, there should not be large differences between measures of our dependent variables in raw versus normed (standard) scores. This is because, by design, we compare 5-year-olds at one point in time (e.g., 1986) to 5-year-olds at a second point in time (e.g., 1988) to 5-year-olds at a third point in time (e.g., 1990), and so on (and then repeat this design structure with many other ages as well). The norming process adjusts for age differences, which is not especially relevant when same-age respondents are being compared. For reasons that are not documented, there are a few raw scores in the NLSY that were not adjusted by the norming pattern; for some of the PIAT assessments, for example, there are very slightly larger sample sizes for the raw scores than for the standard scores. Thus, although the correlations are not exactly unity, as might be expected, in fact, the differences between the normed and raw cognitive ability scores in our study are empirically trivial. Correlations between the two sets of scores across ages and assessments are typically above .90 and often as high as .99. Previous Flynn Effect studies that used across-age designs have found different results when normed and raw scores are used. Although this difference is critical in those designs, it should not lead to important differences in the context of our design. We will verify this empirically, at least in part as a validity check.

We will present both graphical results and regression analyses to test for the Flynn Effect. When we regress the test scores on year, past Flynn Effect findings predict an unstandardized regression coefficient of approximately .33. Further, maternal IQ and the maternal age variables can be entered as covariates in these regression models to linearly control for the threat to validity – the selection bias – described above. (We note that the .33 predicted increase is actually a prediction within the overall IQ metric; because the fluid domain in the past has carried all of the Flynn Effect within it, a relatively pure measure of fluid intelligence like PIAT-M may show a bigger gain than .33 per year, to compensate for no effect in crystallized intelligence.)

3. Results

Normed scores for PIAT-M, PIAT-RR, PIAT-RC, PPVT, and Digit Span were weighted using sampling weights, and means and variances computed for the relevant age groups (for age categories in which $n>30$ for at least 6 consecutive years). In Fig. 1a–e, we present means for the normed (standard) assessment scores. All showed a generally increasing mean pattern over time. These increases could be caused by the Flynn Effect, or by the maternal IQ selection bias discussed above.

Before using mother's IQ, we empirically addressed the two issues of whether mother's IQ can be used as an effective covariate. The first concern was whether mother's IQ itself shows a Flynn Effect, as well as accounting for the selection bias discussed in Methods. In fact, mother's IQ in the NLSY actually has a positive correlation with mother's age in 1979 (when the study began), rather than the negative correlation that the Flynn Effect would predict. For the overall female sample (without regard for having had children), the correlation between mothers' IQ and age-in-1979 was $r=.19$. This small positive correlation is probably caused by the fact that the AFQT scores were not normed within age. In

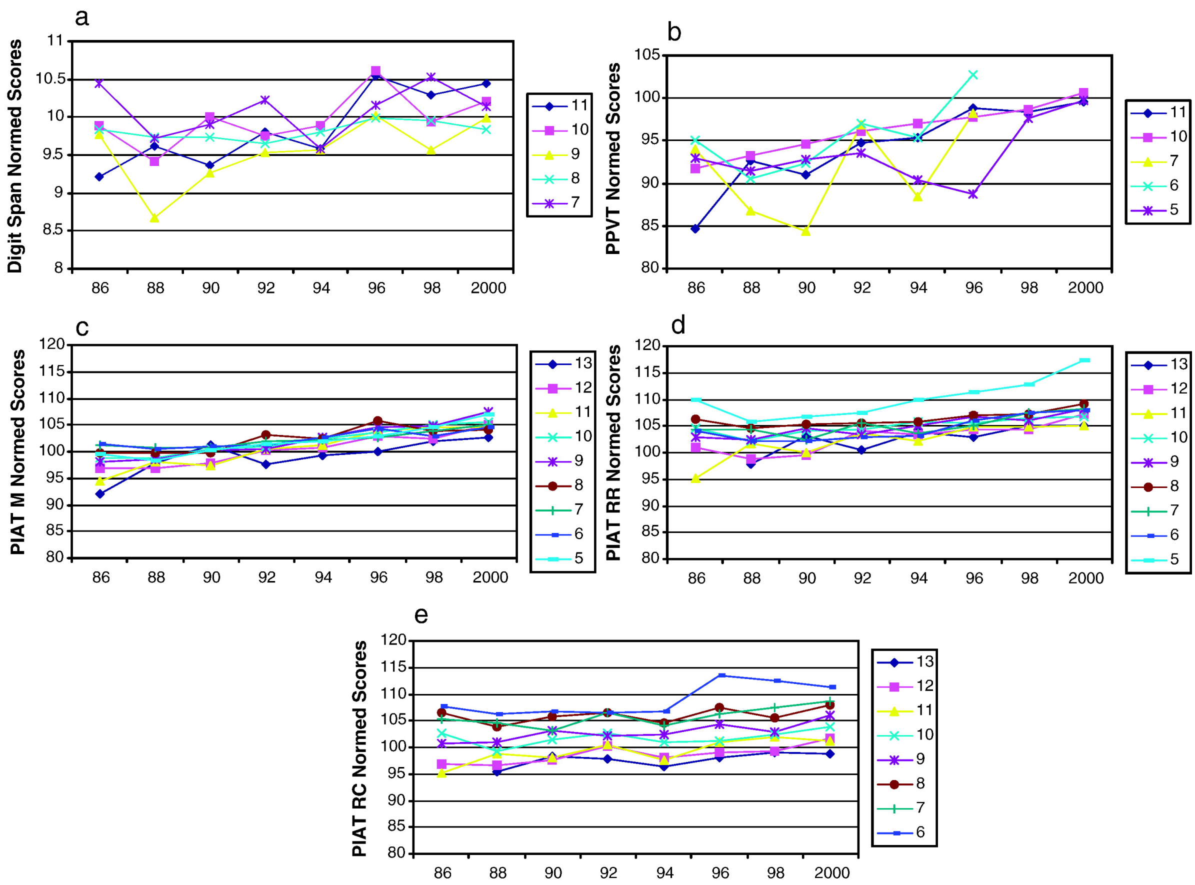

Fig. 1. (a) Digit Span normed score means for ages 5–11 during 1986–2000; (b) PPVT normed score means for ages 5–11 during 1986–2000; (c) PIAT-M normed score means for ages 5–13 during 1986–2000; (d) PIAT-RR normed score means for ages 5–13 during 1986–2000; (e) PIAT-RC normed score means for ages 6–13 during 1986–2000.

1980, when these IQ scores were obtained from the (eventual) NLSY mothers, the NLSY females ranged in age from 15 to 23; it seems reasonable that 15-year-olds would not do as well as 23-year-olds on such a test, because they have not reached the adult asymptote in intellectual growth.

However, in the context of the overall model, when mothers' age-in-1979 was added to the model for each age in the context of our NLSYC design, it never contributed significantly to the predictability. In other words, the positive correlation – explainable through the use of adult age norms for adolescents who were not quite adults – explained no additional variance beyond that explained by mother's IQ and the variable accounting for the Flynn Effect. Thus, it does not appear that the mother's IQ itself needs to be adjusted by maternal age, nor is there any type of Flynn Effect in mother's IQ that is potentially influencing the NLSYC scores by absorbing some of the Flynn Effect in those scores. We note that there might well be a Flynn Effect underlying these scores, masked by the bias caused by not using age-normed scores. This effect would be expected to be relatively small, since there is a range of only seven years in the age of the mothers, which would imply a Flynn Effect of around 2+ IQ points. However, this concern is moot with respect to the current study. We are not investigating whether there is a Flynn Effect for the mothers, but rather intend to use their scores as a statistical adjustment to handle the well-known selection bias in using the NLSYC for this type of design. What we have found empirically is that there is not a negative correlation across ages for the mothers that would absorb part of the potential Flynn Effect for their children.

The second concern with using mother's IQ is whether its correlation with our dependent variables – the child cognitive ability measures – places it within the range of appropriate and valuable covariates. If the correlation is too high, we may covary out useful structural variance. If it is too low, it will not be an effective linear adjustment. Ultimately, of course, this is an empirical question that will be addressed in the context of fitting linear models. But the raw correlations support its utility. We observed the mother's IQ relationship with the child's PIAT-Math scores for each of the age-fixed design structures described in Methods with the following correlations: the correlations ranged from r=.38 for 5-year-olds to r=.44 for both 9- and 10-year-olds, a relatively small range indicating a stable correlation across time. Correlations for other cognitive assessments were similar. These are significant and moderate correlations, as expected, and suggest that a linear adjustment for mother's IQ can have the effect that we wish it to — to adjust for the selection bias. However, in each case, less than 20% of the variance in child's IQ is related to differences in mother's IQ, which argues against the degree of colinearity that would create concern over covarying out critical variance in our DVs. As a result, it appears that we are left with the mother's IQ score as a relatively uncontaminated adjustment for the selection bias, which was the goal of using the mother's IQ score in the first place.

For each age, we defined a regression model regressing normed (standard) test scores on year, separately for each age; these models estimated the slopes and intercepts underlying the mean patterns in Fig. 1. In a second set of regressions, we added mother's age and mother's IQ as covariates; as noted above, mother's IQ was always a significant covariate, whereas mother's age was not, and

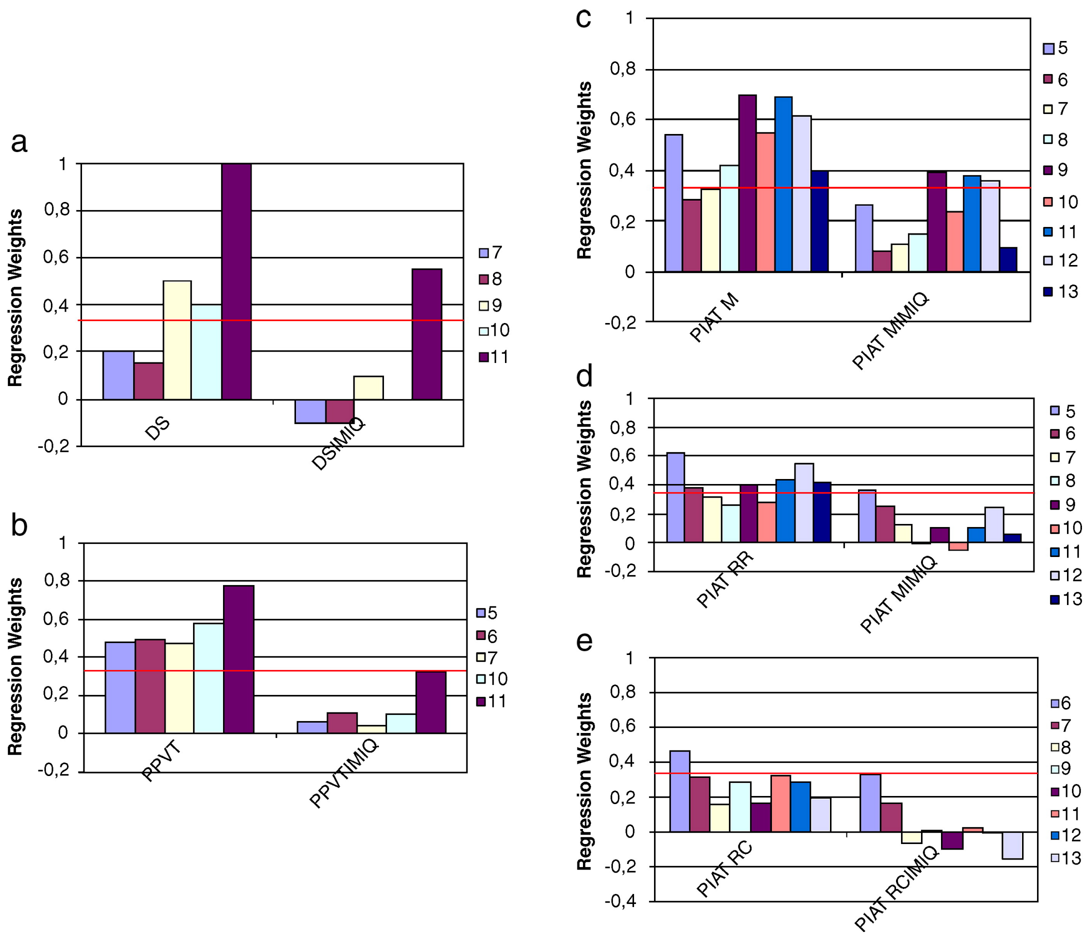

Fig. 2. (a) Regression coefficients (unstandardized b's) from regressing normed test score on year, Digit Span (ignoring on left/controlling on right for mom's IQ); (b) regression coefficients (unstandardized b's) from regressing normed test score on year, PPVT (ignoring on left/controlling on right for mom's IQ); (c) regression coefficients (unstandardized b's) from regressing normed test score on year, PIAT-M (ignoring on left/controlling on right for mom's IQ); (d) regression coefficients (unstandardized b's) from regressing normed test score on year, PIAT-RR (ignoring on left/controlling on right for mom's IQ); (e) regression coefficients (unstandardized b's) from regressing normed test score on year, PIAT-RC (ignoring on left/controlling on right for mom's IQ).

so mother's age was dropped out of these models. The slopes in these models estimate the mean change per year in the cognitive ability. The slopes for Digit Span – which has a normed mean of 10, compared to the other normed means of 100 – were multiplied by 10 to equate the Digit Span means to the means from the other assessments.

In the unadjusted regressions of normed scores, virtually all of the age category slopes were significant: 4 of 7 for Digit Span, 5 of 5 for PPVT, 8 of 8 for PIAT-RC, 9 of 9 for PIAT-RR, and 9 of 9 for PIAT-M. When mother's IQ was included, few of the results stayed significant: 1 of 7 for Digit Span (mean slope per year=.03), 2 of 5 for PPVT (mean slope per year=.12), 2 of 8 for PIAT-RC (mean slope per year=.02), and 4 of 9 for PIAT-RR (mean slope per year=.13). However, 7 of 9 of the PIAT-M slopes remained significant. The mean slope per year for PIAT-M was .23; for the oldest five ages, which are measured more reliably, the mean slope per year was .30 (median=.36). Fig. 2a–e shows those regression slopes for each assessment, with a line drawn in these figures at a slope=.33 (the slope expected based on previous Flynn Effect research; note, however, the comment above suggesting that this expectation might be higher for PIAT-M).

In the equivalent analyses using raw scores instead of normed (standard) scores, the pattern of findings was almost identical. In the unadjusted regressions, the age category slopes were typically significant, as before: 6 of 7 for Digit Span, 5 of 5 for PPVT, 8 of 8 for PIAT-RC, 9 of 9 for PIAT-RR, and 9 of 9 for PIAT-M. When mother's IQ was included in the model, most of the *b*'s became non-significant; 2 of 7 were still significant for Digit Span (mean slope per year=.01), 2 of 5 for PPVT (mean slope per year=.14), 1 of 8 for PIAT-RC (mean slope per year=.00), and 7 of 9 for PIAT-RR (which might be suggestive of some potential remaining Flynn Effect, although the mean slope per year=.08). For these crystallized measures, there did not seem to be a meaningful difference for the normed score and raw score analysis; three of the mean slopes were smaller in the standard score analysis, one was larger.

Table 1
Unstandardized regression coefficients predicting PIAT-Math scores from year (replicated across nine ages) for both raw PIAT-M scores and standard (normed) PIAT-M scores; these coefficients are interpreted as "change per year"

Age	Raw score *b*	Raw score *b* with MIQ adjustment	Standard score *b*	Standard score *b* with MIQ adjustment
5	.15	.06	.55	.28
6	.20	.09	.30	.08
7	.28	.11	.33	.11
8	.31	.12	.42	.13
9	.46	.26	.69	.40
10	.35	.14	.57	.23
11	.48	.26	.67	.38
12	.46	.26	.62	.36
13	.53	.31	.40	.10

All coefficients in this table are statistically significant.

All of the PIAT-M raw score coefficients (9 of 9) were statistically significant. In general, the size of the raw score coefficients indicating change per year was slightly smaller than for the normed score coefficients. To give a sense of this comparison, we present the coefficients under the two procedures for PIAT-M in Table 1. For PIAT-M, the mean slope per year was .21 (compared to .23 for the normed scores); for the oldest five ages, the mean slope was .25 (compared to .30 for the normed scores).

4. Discussion

In the NLSY, we identified a consistent and significant Flynn Effect of approximately the predicted magnitude in PIAT-M, the ability domain for which it was predicted. In most other cases, the Flynn Effects identified for other cognitive ability domains disappeared when we used mother's IQ to (at least partially) adjust for the selection bias caused by women who are younger when they have children themselves having lower IQ's. Once mother's IQ is controlled, we find a generally statistically flat slope for the cognitive assessments related to crystallized intelligence; most of these show an average small positive, non-significant increase across the 14-year period.

We have ruled out that this effect is contaminated by a similar trend in the mother's IQ; rather, the positive correlation between mother's IQ and their age-in-1979 suggests higher IQ scores from older mothers. The explanation for this finding is that the younger members of the original NLSY sample had not reached the age at which adult norms would be fully valid, and their IQs are probably under-estimated. The findings themselves rule out the possibility of an artifact in the other direction, since the cognitive measures show consistently higher secular trends before controlling for mother's IQ than afterwards.

The Flynn Effect has been identified in a number of different data sources, but never in a national probability sample with the kind of broad information available in the NLSY. Chase-Lansdale et al. (1991) document the value of the NLSYC to research efforts within Psychology; this data source has supported thousands of published research studies in the broad social/ behavioral science literature. (These public-domain data may be downloaded from a BLS website, http://www.bls.gov/nls). NLSYC variables include measures of the home environment, problem behaviors/delinquency, alcohol and drug use, social development including dating and sexuality, family income, school information, and extensive information about siblings and parents; further, many of these measures are available with longitudinal structure (though the adolescent-onset behaviors are not collected until age 15 or after; this older subset of the NLSYC data is called the NLSY-Young Adult dataset). Wicherts et al. (2004), among others, note the importance of using longitudinal data with different ages in evaluating the Flynn Effect. Given a Flynn Effect identified within the NLSY, dozens of important questions can be addressed. Specifically, this finding sets the stage for investigating many boundary-condition questions. We conclude by identifying a number of such questions.

First, there are a number of demographic comparisons that are important. Does the Flynn Effect occur across race subgroups (for past discussion, see Ceci, Rosenblum, & Kumpf, 1998; Rushton, 2000)? Minority respondents were oversampled in the NLSY survey design, providing extra precision to address this question. Is the effect the same for males and females? For children from high SES and low SES families? For children with high IQ mothers versus low IQ mothers? For children in urban versus rural settings? For children from different U.S. geographic regions?

Second, important psychometric and item-level issues can be addressed. We identified a Flynn Effect in both normed and raw scores (which are very similar measures in a study with fixed ages), of the type used in past Flynn Effect research. We note that the Flynn (1987) findings show a Flynn Effect without respect to issues of re-norming. Further, Flynn (1987) and later studies found larger IQ gains on raw scores than on normed scores (whereas we found very slightly smaller effects for the raw scores, though in the context of a different type of design). We are hesitant to interpret these as providing a highly valid comparison of raw versus normed scores, because of our fixed age design. Of perhaps of even greater value, individual item-level scores are available for the assessments within the NLSY. Evaluating which of the PIAT-M items contributed most importantly to the Flynn Effect pattern identified here – and which ones did not – would be a useful project. Further, there exist items in the other assessments that partly load on fluid intelligence; can a Flynn Effect be identified for these? Is the appearance of some residual Flynn Effect in PIAT-RR related to items on this scale that may reflect fluid rather than crystallized intelligence? Next, the relation between the Flynn Effect and *g* – general intelligence – has received attention, but needs more focus (see Neisser, 1998, p. 9 for discussion; also see Rushton, 1999). Which part of the PIAT-Math ability distributions caused the NLSY Flynn Effect? What other variables correlate with the increase over age that we observed in the size of the PIAT-Math Flynn Effect?

Third, many variables useful to evaluating the social theories suggested by past researchers are available in the NLSY. Information about sibling and peer interactions is available, along with information about adolescent social behaviors (e.g., dating) and measures of social conflict. Fourth, parenting variables are available. Do the number of trips to the museum, the number of books and magazines, or the frequency of parental punishment/intervention affect the Flynn Effect results? Fifth, the siblings in the NLSY have been distinguished as twins, full siblings, and half siblings (e.g., Rodgers et al., 1994). As a result, biometrical designs and analyses can be conducted to separate genetic/biological variance from shared environmental variance. Sixth, using information about family background and migration, the heterosis hypothesis (Mingroni, 2004) can be evaluated. This hypothesis makes specific predictions about cognitive ability in relation to migration patterns, even across relatively short time periods.

Finally, there are undoubtedly other Flynn Effect questions that could be addressed using the NLSYC that we have not anticipated. The identification of a Flynn Effect in the PIAT-M – and its absence in the other scales – opens up broad potential for addressing meaning, causes, and boundary conditions in future Flynn Effect research.

Acknowledgement

The authors express their appreciation to James Flynn, Pete Legre, and one other anonymous reviewer, whose insightful suggestions improved this paper.

References

Blair, C., Gmson, D., Thorne, S., & Baker, D. (2005). Rising mean IQ: Cognitive demand of mathematics education for young children, population exposure to formal schooling, and the neurobiology of the prefrontal cortex. *Intelligence*, *33*, 93−106.

Brand, C. (1996). G, genes, and pedagogy: A reply to seven (lamentable) chapters. In D. Detterman (Ed.), *Current topics in human intelligence. The environment*, *Vol. 5* (pp. 113−120). Norwood NJ: Ablex.

Ceci, S. J., Rosenblum, T. B., & Kumpf, M. (1998). The shrinking gap between high- and low-scoring groups: Current trends and possible causes. In Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 287−302). Washington, DC: American Psychological Association.

Chase-Lansdale, P. L., Mott, F. L., Brooks-Gunn, J., & Phillips, D. A. (1991). Children of the National Longitudinal Survey of Youth: A unique research opportunity. *Developmental Psychology*, *27*, 918−931.

Colom, R., Lluis-Font, J. M., & Andres-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, D. (2003). IQ on the rise: The Flynn Effect in rural Kenyan children. *Psychological Science*, *14*, 215−219.

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.

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. (1996). What environmental factors affect intelligence: The relevance of IQ gains over time. In D. Detterman (Ed.), *Current topics in human intelligenceThe environment*, *Vol. 5* (pp. 17−30). Norwood, NJ: Ablex.

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

Flynn, J. R. (2006). Efeito Flynn: Repensando a inteligência e seus efeitos [The Flynn Effect: Rethinking intelligence and what affects it]. In C. Flores-Mendoza, & R. Colom (Eds.), *Introdução à Psicologia das Diferenças Individuais [Introduction to the psychology of individual differences]* (pp. 387−411). Porto Alegre: ArtMed.

Greenfield, P. M. (1998). The cultural evaluation of IQ. In Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 81−123). Washington, DC: American Psychological Association.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology*, *57*, 253−270.

Jensen, A. R. (1991). Speed of elementary cognitive processes: a chronometric anchor for psychometric tests of g. *Psychological Test Bulletin*, *4*, 59−70.

Jensen, A. R. (1996). Secular trends in IQ: Additional hypotheses. In D. Detterman (Ed.), *Current topics in human intelligence The environment*, *Vol. 5* (pp. 147−150). Norwood, NJ: Ablex.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Kanaya, T., Ceci, S. J., & Scullin, M. H. (2005). Age differences within secular IQ trends: An individual growth modeling approach. *Intelligence*, *33*, 613−621.

Loehlin, J. (1989). Partitioning environmental and enetic contributions to behavioral development. *American Psychologist*, *44*, 1285−1292.

Loehlin, J. (1996). Environment and intelligence: A comment. In D. Detterman (Ed.), *Current topics in human intelligence The environment*, *Vol. 5* (pp. 151−156). Norwood, NJ: Ablex.

Loehlin, J. C. (2002). The IQ paradox: Resolved? Still an open question. *Psychological Review*, *109*, 759−763.

Lynn, R. (1982). IQ in Japan and the United States shows a growing disparity. *Nature*, *306*, 291−292.

Lynn, R. (1990). Differential rates of secular increase of five major primary abilities. *Social Biology*, *37*, 137−141.

Lynn, R. (1998). In support of nutrition theory. In Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 207−218). Washington, DC: American Psychological Association.

Lynn, R., & Hampson, S. (1986). Intellectual abilities of Japanese children: An assessment of 2 ½–8 ½-year-olds derived from the McCarthy Scales of Children's Abilities. *Intelligence*, *10*, 41−58.

Mahlberg, A. (1997). The rise in IQ scores. *American Psychologist*, *52*, 71.

Martorell, R. (1998). Nutrition and the worldwide rise in IQ scores. In Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 183−206). Washington, DC: American Psychological Association.

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

Neiss, M., Rowe, D. C., & Rodgers, J. L. (2002). Does education mediate the relationship between IQ and age of first birth? A behavioural genetic analysis. *Journal of Biosocial Science*, *34*, 259−275.

Neisser, U. (1998). Introduction: Rising test scores and what they mean. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 3−24). Washington, DC: American Psychological Association.

Retherford, R. D., & Sewell, W. H. (1989). How intelligence affects fertility. *Intelligence*, *13*, 169−185.

Rodgers, J. L. (1998). A critique of the Flynn Effect: Massive IQ gains, methodological 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.

Rodgers, J. L., & Rowe, D. C. (1988). The influence of siblings on adolescent sexual behavior. *Developmental Psychology*, *24*, 722−728.

Rodgers, J. L., Rowe, D. C., & May, K. (1994). DF analysis of NLSY IQ/achievement data: Nonshared environmental influences. *Intelligence*, *19*, 157−177.

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.

Rushton, J. P. (1997). Review essay. *Society*, *34*, 78−82.

Rushton, J. P. (1999). Secular gains in IQ not related to the g factor and inbreeding depression — unlike Black–White differences: A reply to Flynn. *Personality and Individual Differences*, *26*, 381−389.

Rushton, J. P. (2000). Flynn effects not genetic and unrelated to race differences. *American Psychologist*, *55*, 542−543.

Smith, S. (1942). Language and nonverbal test performance of racial groups in Honolulu before and after a 14-year interval. *Journal of General Psychology*, *26*, 51−92.

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*, *33*, 349−362.

Teasdale, T. W., & Berliner, P. (1991). Experience of kindergartens in relation to educational level and cognitive abilities in adulthood: A geographical analysis. *Scandinavian Journal of Psychology*, *32*, 336−343.

Teasdale, T. W., & Owen, D. R. (1987). National secular trends in intelligence: A twenty-year cross-sectional study. *Nature*, *325*, 119−121.

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. T. (2000). Forty year secular trends in cognitive ability. *Intelligence*, *28*, 115−120.

Tuddenham, R. D. (1948). Soldier intelligence in World Wars I and II. *American Psychologist*, *3*, 54−56.

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 invariance over time? Investigating the nature of the Flynn Effect. *Intelligence*, *32*, 509−537.

Wichman, A., Rodgers, J. L., & MacCallum, R. C. (2006). A multilevel approach to the relationship between birth order and IQ. *Personality and Social Psychology Bulletin*, *32*, 117−127.