Intelligence 41 (2013) 808–816

Contents lists available at ScienceDirect

Intelligence

Are cognitive differences between countries diminishing? Evidence from TIMSS and PISA

Gerhard Meisenberg [a,*], Michael A. Woodley [b]

[a] Ross University Medical School, Department of Biochemistry, Dominica
[b] Umeå University, Department of Psychology, Sweden

ARTICLE INFO

Article history:
Received 19 November 2012
Received in revised form 18 February 2013
Accepted 29 March 2013
Available online 16 May 2013

Keywords:
Scholastic achievement
TIMSS
PISA
Trend analysis

ABSTRACT

Cognitive ability differences between countries can be large, with average IQs ranging from approximately 70 in sub-Saharan Africa to 105 in the countries of north-east Asia. A likely reason for the great magnitude of these differences is the Flynn effect, which massively raised average IQs in economically advanced countries during the 20th century. The present study tests the prediction that international IQ differences are diminishing again because substantial Flynn effects are now under way in the less developed "low-IQ countries" while intelligence is stagnating in the economically advanced "high-IQ countries." The hypothesis is examined with two periodically administered scholastic assessment programs. TIMSS has tested 8th-grade students periodically between 1995 and 2011 in mathematics and science, and PISA has administered tests of mathematics, science and reading between 2000 and 2009. In both TIMSS and PISA, low-scoring countries tend to show a rising trend relative to higher-scoring countries. Despite the short time series of only 9 and 16 years, the results indicate that differences between high-scoring and low-scoring countries are diminishing on these scholastic achievement tests. The results support the prediction that through a combination of substantial Flynn effects in low-scoring countries and diminished (or even negative) Flynn effects in high-scoring countries, cognitive differences between countries are getting smaller on a worldwide scale.

1. Introduction

Although rising intelligence trends had been described in the early to mid-20th century (e.g., Loehlin, Lindzey, & Spuhler, 1975, pp. 135–139; Tuddenham, 1948), the pervasiveness of the phenomenon was recognized only during the 1980s (Flynn, 1984a, 1987; Lynn & Hampson, 1986), at least for economically advanced Western societies and Japan. At that time the strongest secular gains were observed in Continental Europe and Japan, possibly as part of their recovery from the setbacks of World War II. This seemingly ubiquitous trend became known as the Flynn effect (Herrnstein & Murray, 1994), or the Lynn–Flynn effect (Rushton, 1999). Table 1 presents some examples of Flynn effects in Western countries.

More recent studies show that in some European countries, the Flynn effect seems to have ended or even gone in reverse (Shayer & Ginsburg, 2009; Sundet, Barlaug, & Torjussen, 2004) although intelligence may still be rising in the United States (Ang, Rodgers, & Wänström, 2010; Flynn & Weiss, 2007). The studies of Teasdale and Owen (2005, 2008) with military conscripts in Denmark are perhaps the most detailed time series available for any country (see Table 1). In Britain, Flynn effects on Raven's Progressive Matrices continued to the most recent cohorts for children, but were no longer present in teenagers aged 15–16 years (Flynn, 2009; see Table 1). This example shows that robust Flynn effects in children do not always predict higher intelligence of adolescents and adults. They can represent accelerated childhood development without higher final intelligence.

* Corresponding author.
E-mail addresses: Gmeisenberg@rossmed.edu.dm (G. Meisenberg), Michael.Woodley@psy.umu.se (M.A. Woodley).

0160-2896/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.intell.2013.03.009

While Flynn effect gains are becoming small, erratic or negative in modern Western societies, rising IQ trends have recently been observed in some non-Western societies. These are summarized in Table 2. Taken together, the results suggest that cognitive disparities between advanced industrial societies and the less developed countries are diminishing today. Because of the commonly observed positive relationship between IQ (or IQ gains) and economic growth (Meisenberg & Lynn, 2012; Weede & Kämpf, 2002), this trend would most likely be associated with diminishing economic inequalities between countries.

A systematic test of this hypothesis requires a set of nationally representative cognitive test data that are available for an appreciable number of countries at different levels of economic and cognitive development. Because of the inaccuracies that are introduced by sampling error and the relatively modest magnitudes of most reported Flynn effects (Tables 1 and 2), longitudinal assessments of Flynn effects require test administrations that are separated by at least 8 to 10 years, although far longer intervals of several decades are desirable. The only data sets of this kind that are available at this time are the results of international assessments of scholastic achievement in mathematics, science and reading. Correlations of pooled scholastic assessment scores with IQ are approximately $r = .90$, at least for countries with high data quality (Lynn & Meisenberg, 2010; Meisenberg & Lynn, 2011). The relationship is so close that both types of test can be considered measures of “intelligence.”

2. Methods

2.1. Properties of TIMSS and PISA

So far only two international scholastic assessment programs have produced trend data over a (marginally) sufficient length of time. These are TIMSS (Trends in International Mathematics and Science Study, formerly Third International Mathematics and Science Study) and PISA (Programme of International Student Assessment). TIMSS tests children in 4th and 8th grade in mathematics and science, usually with greater numbers of countries participating in the 8th grade assessments, which include students with an average age of at least 13.5 years. Assessments began in 1995, and have been repeated at 4-year intervals until 2011.

PISA examines “students who are aged between 15 years 3 months and 16 years 2 months at the time of the assessment, and who have completed at least 6 years of formal schooling” (OECD, 2010, p. 24). The tests cover mathematics, science and reading, with different emphasis in different years. PISA has administered tests every 3 years since 2000. The last published results are from 2009. For the 49 countries having information about TIMSS, PISA and IQ, the correlation between the averaged TIMSS scores and the averaged PISA scores is $r = .876$; between TIMSS and the most recent update of national IQs (Meisenberg & Lynn, 2011), $r = .851$; and between PISA and IQ, $r = .908$.

Both TIMSS and PISA are graded with methods of item response theory, which avoid non-linear relationships between scores and latent proficiency. Both are scaled on a 500/100 scale. In TIMSS, the measurement scale was established based on the 1995 assessment. In later years, the same scale was maintained based on the performance on the preceding assessment of those countries that participated in both. The actual within-country standard deviation, reported separately for mathematics and science, averages 75 to 80. In PISA, the mean value of 500 for each assessment is defined as the average of the participating OECD countries, and the average OECD within-country standard deviation is set at 100. These scaling methods introduce an important limitation: Because the measurement scale is calibrated by the average performance of a group of countries rather than by absolute standards of performance, it is easy to determine changes in *relative* performance when comparing countries,

Table 1
Some Flynn effect gains in Western countries.

Country	Test	Age	Birth cohort	Gain/decade	Source
New Zealand	Otis	10–13	1923/26–1955/58	2.42	Elley (1969)
USA	Stanford–Binet	6–12	1920/26–1960/66	1.95	Flynn (1984b)
USA	Stanford–Binet	13–18	1914/19–1954/59	2.40	Flynn (1984b)
Netherlands	Raven SPM selection	18	1934–1964	7.03	Flynn (1987)
France	Raven	18	1931–1956	10.05	Flynn (1987)
France	Math/Verbal	18	1931–1956	3.74	Flynn (1987)
France	ECNI	10	1955–1978	2.61	Bradmetz and Mathy (2006)
Norway	Math/Verbal/Matrices	19	1935–1957	4.55	Sundet et al. (2004)
Norway	Math/Verbal/Matrices	19	1957–1983	0.33	Sundet et al. (2004)
Denmark	Børge Prien's Prøve	18	1941–1971	2.70	Teasdale and Owen (2005)
Denmark	Børge Prien's Prøve	18	1970–1980	1.65	Teasdale and Owen (2008)
Denmark	Børge Prien's Prøve	18	1980–1985/86	−1.49	Teasdale and Owen (2008)
Sweden	Block design/memory	35–80	1909–1969	2.18	Rönnlund and Nilsson (2008)
Britain	Raven SPM	8–12	1926/30–1967/71	2.48	Flynn (2009)
Britain	Raven SPM	8–12	1967/71–1996/00	2.00	Flynn (2009)
Britain	Raven SPM	13–15	1923/25–1964/66	1.66	Flynn (2009)
Britain	Raven SPM	13–15	1968/75–1996/03	−0.87	Flynn (2009)

Table 2
Some Flynn effect gains in non-western countries.

Country	Test	Age	Birth cohort	Gain/decade	Source
Japan	WISC	6–10	1941/45–1965/69	6.23	Flynn (1987)
Japan	WISC	11–15	1936/40–1960/64	6.65	Flynn (1987)
South Korea	Several	5–16	1970–1990	7.70	te Nijenhuis, Cho, Murphy, and Lee (2012)
Turkey	Draw-a-Person	11	1966–1999	1.59	Kagitcibasi and Biricik (2011)
Argentina	Raven SPM	13–24	1940/51–1974/85	6.28	Flynn and Rossi-Casé (2012)
Brazil	Draw-a-Person	7–11	1919/23–1991/95	2.36	Colom, Flores-Mendoza, and Abad (2006)
Kenya	Raven CPM	7.4 (avg.)	1976/77–1990/91	8.01	Daley, Whaley, Sigman, Espinosa, and Neumann (2003)
Sudan	WAIS-R full scale	50 (avg.)	1937–1957 (avg.)	2.05	Khaleefa, Sulman, and Lynn (2009)
Sudan	WAIS-R verbal	50 (avg.)	1937–1957 (avg.)	−0.83	Khaleefa et al. (2009)
Sudan	WAIS-R performance	50 (avg.)	1937–1957 (avg.)	3.60	Khaleefa et al. (2009)
Dominica	Raven SPM	18–25/51–62	1978/87–1941/54	5.14	Meisenberg, Lawless, Lambert, and Newton (2005)
Saudi Arabia	Raven SPM	8–18 male	1958/68–1992/02	0.29	Batterjee (2011)
Saudi Arabia	Raven SPM	8–18 female	1958/68–1992/02	2.65	Batterjee (2011)
Taiwan	WISC	6–16	1981/91–1991/01	2.33	Chen, Liao, Chen, Chen, and Lynn (2013)
Taiwan	WAIS	16–84	1917/85–1927/95	0.54	Chen et al. (2013)

but difficult to draw inferences about changes in *absolute* performance from an earlier to a later date.

2.2. Measures used

TIMSS and PISA scores are publicly available. TIMSS scores were obtained from the publications of the TIMSS & PIRLS International Study Center at http://timssandpirls.bc.edu. Of the separately published grade 7 and grade 8 scores in 1995, only the grade 8 scores were used because this more closely matches practice in the later assessments. For PISA 2000 and 2009, the scores were obtained from the publicly available data files at http://pisa2000.acer.edu.au/downloads.php and http://pisa2009.acer.edu.au/downloads.php. For both TIMSS and PISA, the measurement scale was re-calibrated to obtain equal score averages for the sets of countries participating in successive assessments. To obtain a "general" cognitive ability measure, all available subject scores (mathematics and science in TIMSS, and reading, mathematics and science in PISA) were averaged with equal weights.

TIMSS trend and *PISA trend* are defined as the difference between scores on the first and last assessments. For PISA, these were from 2000 and 2009. For TIMSS, the number of countries participating in both the first (1995) and last (2011) assessment (19 countries) was rather small. Therefore the trend measure was calculated separately for the 1995–2011, 1995–2007, and 1999–2011 periods. A composite of these three measures, available for 36 countries, was used as the TIMSS trend. Positive values indicate a rising trend.

School enrolment trend for TIMSS was calculated from twoc sources: (1) the Barro–Lee data set published at http://www.barrolee.com, using the change in secondary school enrolment and total years of schooling for the 15–19 years age group between 1995 and 2010; and (2) the most recent data on the net secondary school enrolment ratio from the Human Development Reports of the United Nations. Positive values indicate rising enrolment. This should be considered a very crude measure for changes in the proportion of children who are still in school in grade 4. For PISA, information about the proportion of 15-year-olds who are still in school is provided in OECD (2003), pp. 249–254 and OECD (2010), p. 171–176. The difference between the figures in these two sources was taken as the enrolment trend for PISA.

Economic growth is the ratio of log-transformed per-capita GDP from 5 years before the last to 5 years before the first assessment, with positive values indicating rising prosperity. GDP is obtained from the Penn World Tables (Heston, Summers, & Aten, 2009).

Maternal education trend is operationalized as the difference in average years of female education in the 30–34 years age bracket between the 1980–1990 and 1995–2005 periods. Data are from the Barro–Lee dataset. Positive values indicate a rising trend.

Differential fertility is the partial correlation between education and number of children (sex and age controlled) from the World Values Survey (WVS) Official Aggregate v.20090901, 2009, available at www.worldvaluessurvey.org. Data are based on interviews conducted between 1981 and 2008 with a total of 355,298 respondents in 96 countries and territories, including most of the countries represented in TIMSS and PISA. Data from the age cohorts 17–41 (females) and 19–45 (males) were used because these cohorts comprise, approximately, the generation of the parents of the TIMSS and PISA students. Positive values mean that highly educated persons have more children than the less educated.

These additional variables were included based on the following hypotheses: (1) Everything else being equal, rising school enrolment will lower performance on the scholastic assessments. This is expected because in most countries, low-performing students leave school earlier than those with higher performance, and increased enrolment can be achieved by keeping more low-performing children in school. The scholastic assessments measure only the cognitive level of school children and are therefore not representative of all children. (2) Greater prosperity is expected to create enriched environments for children, raising their intelligence. Therefore cognitive skills are expected to rise in countries in which general prosperity rises. (3) Higher maternal education is expected to favor children's intelligence because more educated mothers are expected to create more favorable environments for their children's cognitive development. Therefore children's intelligence is expected to rise in countries in which female education has increased in the mothers' generation. (4) Differential fertility by education affects the proportion of children who are raised by highly educated versus less educated parents. Through genetic and environmental mechanisms, children of highly educated parents are expected to be more likely than those of less educated parents to have high intelligence and do well in school. Therefore rising intelligence and scholastic achievement are predicted for countries in which highly educated parents have more children than those with less education. For example, the expected positive effect of rising female education on children's intelligence will not materialize if highly educated women remain childless.

3. Results

3.1. Levels and trends in performance

PISA has been designed primarily for the OECD countries. Therefore most results are for economically advanced countries, and variance in scores is low. The trend measure is available for the 42 countries that participated in both the first and the last assessment, conducted in 2000 and 2009, respectively. With the scaling procedure outlined under Methods, mean and standard deviation in 2000 were 476.1 ± 57.4 with a range from 315.8 (Peru) to 556.8 (Netherlands); and in 2009, mean and standard deviation were 481.3 ± 46.1, ranging from 369.9 (Peru) to 546.5 (Hong Kong). The lower scoring and economically less developed countries in the sample are Argentina, Brazil, Chile, Mexico, Peru, Indonesia and Thailand. When results were scaled to equal means for the 2000 and 2009 assessments, the average absolute difference between 2000 scores and 2009 scores for the 42 countries was 16.4 points (approximately 2.46 IQ points, or 2.73 IQ points/decade). Changes ranged from −35.6 (5.34 IQ points) in the United Kingdom to +48.9 (7.34 IQ points) in Peru.

Like PISA, TIMSS includes mainly economically advanced countries, but there is a somewhat better representation of lower-scoring, economically less developed countries. The lowest scoring countries (average of first and last scores) are Morocco (366.0), Colombia (407.3) and Kuwait (407.5), and the highest scoring countries are Singapore (619.7), Taiwan (596.9) and South Korea (593.1). Mean and standard deviation for the 36 countries were 505.8 ± 60.2 in 1995/99, and 501.6 ± 58.2 in 2007/11. When results are scaled to equal means for the first and last assessments, the average absolute score difference is 20.9 points (3.69 IQ points assuming within-country standard deviation of 85 in TIMSS, or 2.46 points/decade). The steepest declines were observed in Malaysia (−71.5 TIMSS points), Thailand (−67.2), and Bulgaria (−62.9). The steepest rises were in Lithuania (48.9 points), Morocco (44.5 points), and Turkey (37.5 points). For the 22 countries with trend data for both TIMSS and PISA, the correlation between PISA trend and TIMSS trend is non-significant (but positive) at $r = .199$.

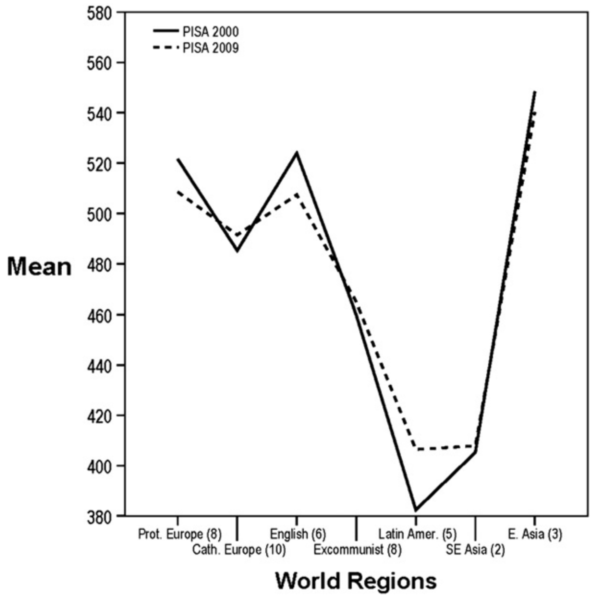

Fig. 1. Profile of PISA scores in different world regions. PISA 2009 averages are compared with PISA 2000 averages. Scores are scaled to the same overall mean for the 2000 and 2009 assessments. Numbers in parentheses indicate the number of countries.

Figs. 1 and 2 show the trends in different world regions. There is little systematic change in the countries of Continental Europe and East Asia. Results from the English-speaking countries are inconsistent, with TIMSS showing gains and PISA losses. All English-speaking countries

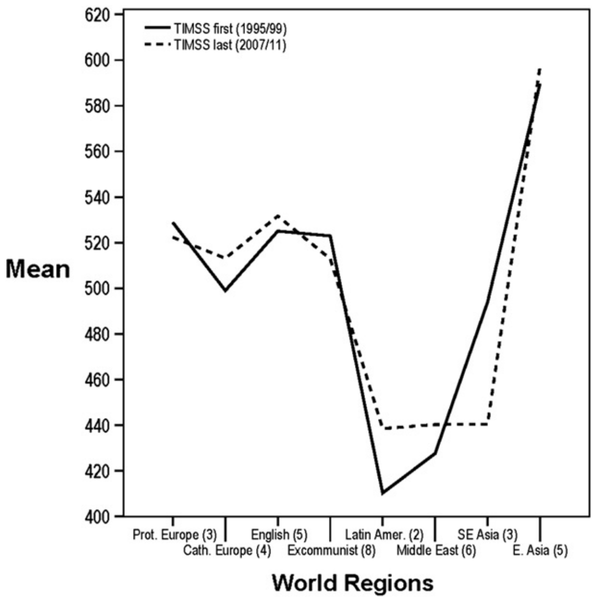

Fig. 2. Profile of TIMSS scores in different world regions. TIMSS 2007/11 averages are compared with TIMSS 1995/99 averages. Scores are scaled to the same overall mean for the first and last assessments. Numbers in parentheses after the world regions indicate the number of countries.

except the United States have declining trends in PISA, but TIMSS finds rising trends for Canada, the United Kingdom and the United States. There may be a marginal rise in the countries of Catholic Europe, and a marginal decline in Protestant Europe. Latin American countries were improving in both TIMSS and PISA except for Argentina, which declined slightly in PISA. Most countries of the Middle East and North Africa have inconsistent trends of modest size, but there are sound gains in Morocco and Turkey. Trends in Southeast Asia are declining for Malaysia and Thailand, but ambiguous for Indonesia. The appendix shows the scores listed by country.

3.2. Predictors of the PISA trend

Table 3 shows the correlations of the PISA trend with hypothesized predictors. As predicted, the correlations between the trend measure and performance level are negative, meaning that low-scoring countries tend to have rising trends relative to high-scoring countries. Inevitably, the correlation between trend and performance level is more negative for performance on the first than the last assessment. We also see that rising performance of children is associated with rising education in the maternal generation as predicted. However, the positive association of rising school enrolment with rising performance is contrary to predictions. Also differential fertility has the "wrong" sign.

However, the enrolment trend is related to the PISA scores. In many low-scoring countries, which are the economically less developed countries, secondary school enrolment has been rising; in high-scoring countries, enrolment increases have been minimal because secondary school enrolment had been near 100% in 2000 already. Also differential fertility is related to test scores, with lower-scoring countries having more negative correlations between parental education and number of children. The latter has been described before. It has been attributed to a more negative education–fertility relationship in countries that have not yet completed the demographic transition (Lynn, 2011; Meisenberg, 2008; Woodley, 2012).

These relationships among predictors necessitate a multivariate analysis. Table 4 shows regression models in which the PISA trend is predicted by the variables in Table 3. Following practice in the economics literature, where economic growth is predicted by *initial* per-capita GDP and additional variables, the first two regressions use *initial* PISA performance as a predictor of the subsequent trend. This practice is justified as long as we assume that differences in test score means and trends are "real," rather than being the result of random measurement error. Models with average instead of initial performance as predictor are nevertheless included, to take account of the latter possibility.

Table 4
Regression models predicting the PISA trend. Standardized β coefficient and t statistic are shown.

	Model 1		Model 2		Model 3		Model 4	
Predictor	β	t	β	t	β	t	β	t
PISA 2000 score	−.937	5.69	−.943	5.73				
PISA 2000/09 avg.					−.755	3.73	−.638	3.90
Enrolment trend	−.294	1.98	−.281	1.90	−.228	1.26	−.178	1.01
Economic growth	.111	1.01			.134	1.03		
Maternal educ. trend	.288	2.41	.284	2.38	.338	2.41	.340	2.48
Differential fertility	.206	1.48	.236	1.74	.133	0.79		
N	40		40		40		41	
Adj. R^2	.552		.552		.379		.376	

The models show that the PISA scores are highly significant negative predictors of the PISA trend in all models, indicating a relative rise of performance in lower-scoring countries. Among the other predictors, rising female education in the parental generation is a consistent though not very strong predictor of rising performance. Rising school enrolment has the expected negative effect, although this does not reach statistical significance in most models. Also differential fertility now has the predicted positive sign, although this predictor does not reach a conventional statistical significance level. Finally, rising prosperity does not seem to have a substantial effect independent of the other predictors.

3.3. Predictors of the TIMSS trend

Table 5 shows the correlations of the TIMSS trend with hypothesized predictors. Only the relationship with initial performance is statistically significant, in the expected direction. Correlations with the earlier and averaged TIMSS scores are again negative, indicating a rising trend for (initially) low-scoring countries. The correlations of the

Table 3
Correlations between PISA trend and hypothesized predictors. N = 40 countries.

Predictor	PISA trend	PISA 2000	PISA 2000–09	PISA 2009	Enrolment trend	Economic growth	Maternal educ. trend
PISA 2000 score	−.684***						
PISA 2000/09 avg. score	−.570***	.989***					
PISA 2009 score	−.409**	.945***	.983***				
Enrolment trend	.357*	−.627***	−.635***	−.625***			
Econ. Growth	.105	.045	.072	.103	−.077		
Maternal educ. trend	.371*	−.199	−.149	−.083	.183	.031	
Differential fertility	−.223	.588***	.617***	.636***	−.410**	.207	.053

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 5
Correlations between TIMSS trend and hypothesized predictors. N = 32 countries.

Predictor	TIMSS trend	TIMSS 1995	TIMSS 1995–07	TIMSS 2007	Enrolment trend	Economic growth	Maternal educ. trend
TIMSS 1995 score	−.389*						
TIMSS 1995–07 avg.	−.162	.971***					
TIMSS 2007 score	.091	.880***	.967***				
Enrolment trend	−.190	−.404*	−.491**	−.553**			
Econ. Growth	−.048	.179	.175	.159	.200		
Maternal educ. trend	.135	.033	.067	.099	.384*	.352*	
Differential fertility	.059	.329	.371*	.392*	−.564**	−.138	−.289

* $p < .05$.
** $p < .01$.
*** $p < .001$.

TIMSS trend with other predictors do not even approach statistical significance. However, similar to observations with PISA in Table 3, we see again that lower-scoring countries tend to have rising school enrolment and a more negative relationship between education and fertility.

The regression models in Table 6 show that when other predictors are included in the model, a rising achievement trend is consistently associated with low initial and average TIMSS scores. Again, lower-scoring countries tend to catch up with higher-scoring countries. Of the remaining four variables, rising school enrolment is associated with declining performance as expected; and as in PISA, rising female education in the parental generation is associated with rising performance. Rising prosperity and differential fertility do not have significant effects.

4. Discussion

4.1. Reality and extent of test score convergence

The narrowing of worldwide cognitive gaps was predicted based on the reasoning that there are biological limits to human intelligence. These limits are being approached in (most of) the higher scoring countries, but not yet in (most of) the lower scoring countries. Therefore the kinds of environmental improvement that have fueled Flynn effects in the recent past are predicted to show diminishing returns in the high-scoring but not the low-scoring countries. This hypothesis predicts that a negative effect of high cognitive performance on subsequent performance trends can be observed even after taking account of measurable changes in environmental conditions that are likely to impact cognitive trends, such as economic growth and better schooling.

Table 6
Regression models predicting the TIMSS trend. Standardized β coefficient and t statistic are shown.

	Model 1		Model 2		Model 3		Model 4	
Predictor	β	t	β	t	β	t	β	t
TIMSS 1995 score	−.670	3.84	−.671	4.16				
TIMSS 1995/07 avg.					−.476	2.19	−.490	2.43
Enrolment trend	−.578	2.87	−.631	3.70	−.554	2.26	−.583	2.75
Economic growth	.065	0.39			.016	0.08		
Maternal educ. trend	.377	2.19	.374	2.47	.385	1.90	.361	2.02
Differential fertility	.072	0.39			.037	0.17		
N	32		35		32		35	
Adj. R^2	.312		.351		.089		.151	

The observation that performance trends in TIMSS and PISA are more positive in countries with lower initial (or even average) performance is compatible with this hypothesis. This is seen in the negative correlations between initial score and the trend measure in Tables 3 and 5. The observation that correlations between the trend measure and the *average* performance on the first and last assessment are still negative in both cases suggests that the relationship is real, rather than being caused entirely by random fluctuations in test scores between the first and the last assessment. Thus we can state with some confidence that overall, lower-scoring countries tend to catch up with the higher-scoring countries.

The magnitude of test score convergence is less certain. In PISA, continuation of the current trends is calculated to erase the differences between high-scoring and low-scoring countries in only 40 years, with a 95% confidence interval of 27 to 77 years. In TIMSS, complete convergence would result after 341 years, with a 95% confidence interval of 70 years to never. These calculations are based on the prediction of the trend measure by the averaged performance on the first and last assessments. We do not know whether test score convergence will ever be complete. It might not if, as is frequently assumed (e.g., Jensen, 1998), biological limits for the development of high intelligence are different in different countries. One possible outcome is partial convergence leading to smaller but persistent gaps, similar to test score convergence between racial groups in the United States during the last three decades of the 20th century (National Center for Education Statistics, 2009).

Because of the relative rather than absolute scaling of the results, it is uncertain whether test score convergence is caused by rising performance in lower-scoring countries, declining performance in higher-scoring countries, or different rates of rise or decline in the two kinds of country. However, the TIMSS organizers describe test score trends for 8th grade mathematics in absolute terms: “Of the 25 countries and eight benchmarking participants with comparable data spanning the 1995 or 1999 to 2011 period, nine countries and four benchmarking participants had increased achievement, eleven countries and two benchmarking participants had decreased achievement, and five countries and two benchmarking participants showed no difference." (Mullis, Martin, Foy, & Arora, 2012, p. 55) For science, the authors counted 9 countries and 3 benchmarking participants with increasing performance, six countries with decreasing performance, and 10 countries and five benchmarking participants with no change (Martin, Mullis, Foy, & Stanco, 2012, p. 53). Thus it appears that in this assortment of countries, most of which are at a rather high level of socio-economic development, there is little systematic change in school achievement at age 14 or 15. This is compatible with the Flynn effects in Tables 1 and 2, which suggest a combination of stagnating IQs in higher-scoring countries with rising performance in lower-scoring countries.

4.2. Other effects on achievement trends

Although performance level was the most robust predictor of temporal trends in both TIMSS and PISA, some other effects of varying strength and consistency were noted. Changes in female education in the generation of the test takers' mothers were the most consistent predictor. This suggests that educational advances and rising cognitive horizons feed back on themselves: raising the educational level of one generation can lead to rising scholastic achievement in the next generation. Much of this is in all likelihood caused by more educated parents providing their children with intellectually more stimulating environments. However, the parental generation is also the generation of the children's teachers. Therefore rising educational levels and rising intelligence in the adult generation can conceivably lead to cognitive gains in children through greater competence of teachers.

Another effect, evident in TIMSS and to some extent in PISA, is a negative effect of rising school enrolment on performance. The likely reason is that in school systems in which many children are no longer in school at age 14 or 15, those who are no longer in school tend to be the less proficient. When enrolment in this age group is raised to near-100%, the inclusion of these less proficient children depresses the average performance of those who are in school. This effect is expected when school systems expand without concomitant improvements in the quality of instruction. For example, Thailand raised the time of compulsory education from 6 to 9 years in 2002. The resulting increase of 8th-grade enrolment is a likely cause for the large performance decline observed in TIMSS (see Appendix).

Finally, the PISA results suggest that differential fertility might play some role. This relationship is theoretically predicted because parental socio-economic status, and especially parental education, is a good predictor of children's intelligence. Therefore everything else being equal, cognitive trends should be influenced by the direction and extent to which childbearing is related to education and intelligence. Although the relationship of fertility with education and social status has been negative in most countries since the mid-19th century (Skirbekk, 2008), and negative relationships between IQ and number of children are observed regularly in modern societies (e.g., Meisenberg & Kaul, 2010), the strength of this relationship varies among countries and is therefore expected to contribute to differences in cognitive trend lines between countries. To date, religiosity is the only psychological trait for which differential fertility has been shown to be an important determinant of temporal trends at the country level (Meisenberg, 2011, 2012).

Importantly, accounting for these additional effects did not eliminate the independent effect of performance level on performance change. Comparison of the raw correlations in Tables 3 and 5 with the β coefficients in Tables 4 and 6 shows that inclusion of the additional predictors strengthens the effect of performance level on the trend measure. The results suggest that in the lower-scoring countries, a more substantial rise in scholastic performance is impeded to some extent by rising secondary school enrolment, which keeps a greater proportion of less capable children in school until the age of testing (14 or 15 years), and possibly by a more negative relationship between education and fertility.

4.3. Relationships between scholastic achievement and intelligence

Despite some issues of sample representativeness, the results of TIMSS and PISA are considered highly accurate measures of student achievement (e.g., Hanushek & Woessmann, 2010). Major effort is invested in each round of assessments, and most country samples in both TIMSS and PISA include approximately 5000 school children. Consequently, these testing programs have spawned an extensive literature that is concerned with the effectiveness of educational systems in individual countries. Most of this literature is about specific curricular subjects such as science, mathematics, or reading literacy.

In marked contrast to this emphasis on the particulars of national school systems and curricular subjects, comparative studies have shown consistently that the determinants of achievement are of a general nature (Rindermann, 2007). For example, at the country level the average correlation between TIMSS mathematics and science scores across all four assessments is $r = .932$, and for PISA mathematics and reading it is $r = .934$. Country-level correlations between averaged scholastic assessment scores and the "national IQs" reported by Lynn and Vanhanen (2006, 2012) are in the vicinity of .90 (Lynn & Meisenberg, 2010; Meisenberg & Lynn, 2011). These high correlations suggest that to a large extent all these tests measure the same construct, which we can label intelligence, or *g*. On the other hand, between-country differences, relative to within-country differences, are 30%–50% greater for scholastic achievement than for IQ (Lynn & Meisenberg, 2010; Meisenberg & Lynn, 2011), most likely because poor quality of schooling in low-IQ countries has a greater effect on scholastic achievement than on IQ. Perhaps one of the reasons for the high correlations between school achievement tests and IQ is that the quality of schooling is determined in large part by the intelligence of the teachers. Thus the scholastic assessments measure not only the intelligence of school children, but to some extent intelligence in the teachers' generation as well. Temporary discrepancies between IQ trends (Flynn effects) and scholastic achievement trends have been noted for the United States and Norway (Flynn, 1987). However, the high correlation between the two types of cognitive test at the country level makes it likely that in most cases, performance on scholastic achievement tests and intelligence tests will change in parallel.

4.4. Limitations of the study

The modern age of periodically administered international scholastic assessments started only in 1995 with the first round of the Third International Mathematics and Science Study (TIMSS, now: Trends in International Mathematics and Science Study). Earlier scholastic assessments, which had been performed since the 1970s, had been limited to a single test administration, and in most cases only "% correct" was reported. Because of poorly reported sample selection and/or non-linear relationships between latent competence and measured test scores, these early scholastic assessments are not suitable for the study of temporal trends. Useful time series are still limited to TIMSS and PISA, which so far span 16 and 9 years, respectively. This is marginal for the assessment of regularities in cognitive trends across countries. Based on the literature (Tables 1 and 2), we can expect that true Flynn effect trends on this time scale will in most cases be less than the equivalent of 4 or 5 IQ points, although larger gains (and losses) are possible. True differences of this magnitude can easily be obscured by sampling artifacts and other biases.

In addition to the marginal time series, the accuracy of the data is questionable. This is evident by the observation that the correlation between TIMSS trend and PISA trend is only .199 for the 22 countries that have trend measures for both, although the two assessment programs cover roughly the same time period (1995–2011 for TIMSS, 2000–2009 for PISA). If the reported achievement scores were highly accurate, this correlation would be higher. Also other variables, including those used for school enrolment, educational changes in the parental generation, and differential fertility are not expected to be very precise. Therefore we should apply the rule that absence of evidence is not evidence of absence. It is likely that future studies with longer time series and more accurate data will reveal relationships that cannot be demonstrated at the present time.

Another limitation of both TIMSS and PISA is the heavy over-representation of prosperous and high-IQ countries. This leaves us with few data points for the most interesting countries, which are the economically less developed countries in which more than half of the world population resides. Although most studies show that Flynn effect gains (or losses) have been small in economically advanced countries after 1995 (Table 1), results from less developed countries are less consistent (Table 2).

4.5. Future prospects

The present study is the first attempt at using TIMSS and PISA for a systematic analysis of global cognitive trends. These two testing programs are ongoing projects that administer tests at regular intervals of 4 and 3 years, respectively. Therefore analysis of longer trend periods will become possible. Also, low-scoring countries that did not participate in the first assessments can increasingly be included.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.intell.2013.03.009.

References

Ang, S. C., Rodgers, J. L., & Wänström, L. (2010). The Flynn effect within subgroups in the U.S.: Gender, race, income, education, and urbanization differences in the NLSY-children data. *Intelligence*, *38*, 367–384.

Batterjee, A. (2011). Intelligence and education: The Saudi case. *Mankind Quarterly*, *52*, 133–190.

Bradmetz, J., & Mathy, F. (2006). An estimate of the Flynn effect: Changes in IQ and subtest gains of 10-yr-old French children between 1965 and 1988. *Psychological Reports*, *99*, 743–746.

Chen, H. -Y., Liao, Y. -K., Chen, Y. -H., Chen, H. -P., & Lynn, R. (2013). Two studies of recent increases of intelligence in Taiwan. *Mankind Quarterly*, *53*, 348–357.

Colom, R., Flores-Mendoza, C. E., & Abad, F. J. (2006). Generational changes on the draw-a-man test: A comparison of Brazilian urban and rural children tested in 1930, 2002 and 2004. *Journal of Biosocial Science*, *35*, 33–39.

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.

Elley, W. B. (1969). Changes in mental ability in New Zealand school children, 1936–1968. *New Zealand Journal of Educational Studies*, *4*, 140–155.

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

Flynn, J. R. (1984b). IQ gains and the Binet decrements. *Journal of Educational Measurement*, *21*, 283–290.

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

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

Flynn, J. R., & Rossi-Casé, L. (2012). IQ gains in Argentina between 1964 and 1998. *Intelligence*, *40*, 145–150.

Flynn, J. R., & Weiss, L. G. (2007). American IQ gains from 1932 to 2002: The WISC subtests and educational progress. *International Journal of Testing*, *7*, 209–224.

Hanushek, E., & Woessmann, L. (2010). Sample selectivity and the validity of international student achievement tests in economic research. *CESifo Working Papers*, *No. 3007*, .

Herrnstein, R. J., & Murray, C. (1994). *The Bell curve: Intelligence and class structure in American life.* New York: Free Press.

Heston, A., Summers, R., & Aten, B. (2009). *Penn World Table version 6.3.* Center for International Comparisons : University of Pennsylvania (available at http://pwt.econ.upenn.edu/php_site/pwt63/pwt63_retrieve.php)

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

Kagitcibasi, C., & Biricik, D. (2011). Generational gains on the Draw-a-Person IQ scores: A three-decade comparison from Turkey. *Intelligence*, *39*, 351–356.

Khaleefa, O., Sulman, A., & Lynn, R. (2009). An increase of intelligence in Sudan, 1987–2007. *Journal of Biosocial Science*, *41*, 279–283.

Loehlin, J. C., Lindzey, G., & Spuhler, J. N. (1975). *Race differences in intelligence.* San Francisco: W. H. Freeman & Co.

Lynn, R. (2011). *Dysgenics. Genetic deterioration in modern populations* (2nd revised ed.). London: Ulster Institute.

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

Lynn, R., & Meisenberg, G. (2010). National IQs calculated and validated for 108 nations. *Intelligence*, *38*, 353–360.

Lynn, R., & Vanhanen, T. (2006). *IQ and global inequality.* Augusta (GA): Washington Summit.

Lynn, R., & Vanhanen, T. (2012). *Intelligence. A unifying construct for the social sciences.* London: Ulster Institute.

Martin, M. O., Mullis, I. V. S., Foy, P., & Stanco, G. M. (2012). *TIMSS 2011 international results in science.* Boston (MA): TIMSS & PIRLS International Study Center.

Meisenberg, G. (2008). How universal is the negative correlation between education and fertility? *Journal of Social Political and Economic Studies*, *33*, 205–227.

Meisenberg, G. (2011). Secularization and desecularization in our time. *Journal of Social, Political and Economic Studies*, *36*, 318–359.

Meisenberg, G. (2012). Demography and modernization as the major determinants of religious belief. *Mankind Quarterly*, *52*, 388–414.

Meisenberg, G., & Kaul, A. (2010). Effects of sex, race, ethnicity and marital status on the relationship between intelligence and fertility. *Mankind Quarterly*, *40*, 152–194.

Meisenberg, G., Lawless, E., Lambert, E., & Newton, A. (2005). The Flynn effect in the Caribbean: Generational change of cognitive test performance in Dominica. *Mankind Quarterly*, *46*, 29–69.

Meisenberg, G., & Lynn, R. (2011). Intelligence: A measure of human capital in nations. *Journal of Social, Political and Economic Studies*, *36*, 421–454.

Meisenberg, G., & Lynn, R. (2012). Cognitive human capital and economic growth: Defining the causal paths. *Journal of Social Political and Economic Studies*, *37*, 141–179.

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics.* Boston (MA): TIMSS & PIRLS International Study Center.

National Center for Education Statistics (2009). *NAEP 2008 trends in academic progress.* NCES 2009-479.

OECD (2003). *Literacy skills for the world of tomorrow. Further results from PISA 2000.* UNESCO Institute for Statistics (Available at www.oecd.org/edu/preschoolandschool/programmeforinternationalstudentassessmentpisa/33690591.pdf)

OECD (2010). *PISA 2009 results: What students know and can do: Student performance in reading, mathematics and science*, *vol. 1*, : OECD Publishing (Available at http://www.oecd.org/pisa/pisaproducts/48852548.pdf)

Rindermann, H. (2007). The big *G*-factor of national cognitive ability. *European Journal of Personality*, *21*, 767–787.

Rönnlund, M., & Nilsson, L. -G. (2008). The magnitude, generality, and determinants of Flynn effects on forms of declarative memory and visuospatial ability: Time-sequential analyses of data from a Swedish cohort study. *Intelligence*, *36*, 192–209.

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.

Shayer, M., & Ginsburg, D. (2009). Thirty years on — A large anti-Flynn effect? (II): 13- and 14-year olds. Piagetian tests of formal operations norms 1976–2006/7. *British Journal of Educational Psychology*, *79*, 409–418.

Skirbekk, V. (2008). Fertility trends by social status. *Demographic Research*, *18* (article 5).

Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, *32*, 349–362.

te Nijenhuis, J., Cho, S. H., Murphy, R., & Lee, K. H. (2012). The Flynn effect in Korea: Large gains. *Personality and Individual Differences*, *53*, 147–151.

Teasdale, T. W., & Owen, D. R. (2005). A long-term rise and recent decline in intelligence test performance: The Flynn effect in reverse. *Personality and Individual Differences*, *39*, 837–843.

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

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

Weede, E., & Kämpf, S. (2002). The impact of intelligence and institutional improvements on economic growth. *Kyklos*, *55*, 361–380.

Woodley, M. A. (2012). The social and scientific temporal correlates of genotypic IQ and the Flynn effect. *Intelligence*, *40*, 189–204.