Intelligence 49 (2015) 1–9

Contents lists available at ScienceDirect

Intelligence

A Cross-Temporal Meta-Analysis of Raven's Progressive Matrices: Age groups and developing versus developed countries

Peera Wongupparaj [a,*], Veena Kumari [a,b], Robin G. Morris [a]

[a] Department of Psychology, Institute of Psychiatry, Psychology, and Neuroscience, King's College London, UK
[b] NIHR Biomedical Research Centre for Mental Health, South London and Maudsley NHS Foundation Trust, London, UK

ARTICLE INFO

Article history:
Received 8 September 2014
Received in revised form 27 November 2014
Accepted 29 November 2014
Available online xxxx

Keywords:
The Flynn effect
Fluid intelligence
Raven's Progressive Matrices
Cross-Temporal Meta-Analysis

ABSTRACT

While many studies have investigated the rise in IQ over time in various countries, the present study attempts to fill the gap in the Flynn effect literature by considering data with various sample sizes, and different study designs, age groups and types of country. A Cross-Temporal Meta-Analysis (CTMA) technique was used to examine the relationship between mean IQ scores from the Raven's Progressive Matrices (RPM) and years of publication, moderated by age group and types of country over a period of 64 years (1950–2014). In all, 202,468 participants were included from 48 countries. We conclude that there is an obvious link between mean IQ scores and years of publication. Importantly, interaction analyses indicate that both age group and types of country moderate this relationship.

Contents

1. Introduction

The term 'Flynn effect' has been coined to designate increases in intelligence scores observed historically in the general population, this phenomenon widely investigated in studies spanning many decades and countries (i.e., Flynn, 1987;

* Corresponding author at: Department of Psychology, P078, Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, London SE5 8AF, UK. Tel.: +44 207 848 5716.
E-mail address: Peera.wongupparaj@kcl.ac.uk (P. Wongupparaj).

http://dx.doi.org/10.1016/j.intell.2014.11.008
0160-2896/

Flynn & Rossi-Casé, 2012; Lynn & Hampson, 1986; Lynn & Harvey, 2008). Although a rise in intelligence scores has been found in the large majority of studies, varying from as small as 1.65 points in Estonia over 72 years to 21.35 points over 34 years in Argentina (Williams, 2013), IQ gains found in specific groups or country comparisons have still been insufficiently investigated.

In general, the majority of studies have employed large cohort techniques to study this phenomenon (Pietschnig, Voracek, & Formann, 2010), their sampling mostly relying on specific criteria, for example, selecting according to age range or certain demographic features, including educational or occupational categories (e.g. school children, college and university students, seniors, conscripts, engineers or twin groups) (Lynn, 2009; Sundet, Barlaug, & Torjussen, 2004a, 2004b; Teasdale & Owen, 2008; Tuddenham, 1948). Therefore, it would be beneficial if the resulting groupings, sample sizes, study designs, or even type of country (developed and developing) are taken into account and analyzed using formal meta-analysis. This attempt may potentially produce a clearer account of the Flynn effect (Rodgers, 1999). Furthermore, a variety of intelligence tests have been used in different countries to study the Flynn effect including, for example, the Wechsler procedures (e.g. China; Liu & Lynn, 2013; Liu, Yang, Li, Chen, & Lynn, 2012), the General Aptitude Test (GATB) (e.g. Holland; Woodley & Meisenberg, 2013), the Coloured Progressive Matrices (CPM) (e.g. Australia and the UK; Cotton et al., 2005; Lynn, 2009) and the Standard Progressive Matrices (SPM) (Saudi Arabia; Batterjee, Khaleefa, Ali, & Lynn, 2013). Accordingly, in order to compare intelligence scores among groups with different demographic features or countries or cultures, the intelligence test should be the same or comparable types of tests used. In this respect, the Raven's Progressive Matrices (RPM), which have three versions, Coloured, Advanced, and Standard Progressive Matrices, may make the best candidate in terms of frequent international use (Van de Vijver, 1997) and construct validity across age, gender, and country (Rushton, Skuy, & Bons, 2003, 2004, Rushton, Skuy, & Fridjhon, 2003). Furthermore, the RPM has been widely used for a long time since the first version of test in 1938, such that it has produced robust data over a sufficiently long time period to conduct a meta-analysis investigating the Flynn effect.

To our knowledge, only a few studies have investigated the Flynn effect in different countries using a cohort method and meta-analyses (Pietschnig et al., 2010; te Nijenhuis, Murphy, & van Eeden, 2011; te Nijenhuis & van der Flier, 2013). In particular, with meta-analyses, less clear results have been found. For example, Brouwers, Van de Vijer, and Van Hemert (2009) conducted meta-analysis on the RPM using published studies dating from 1987 to 2003. Here, the correlation between RPM mean scores and year of publication was negligible, i.e. the Flynn effect was not observed for this sizable dataset (798 samples across 45 countries; $r = 0.07$). Nevertheless, within the sample, the Flynn effect was found in some counties and to varying extents, for instance, with more substantial correlation coefficients found in Iran than in Australia (0.97 *vs* 0.57) and in Poland than in the United States (0.77 *vs* 0.45).

To further investigate the Flynn effect using meta-analysis, this study presents a Cross-Temporal Meta-Analysis (CTMA) of studies that reported data on RPM, CPM, SPM, and APM, comprising samples of children, adults, and older adults from various study designs (quasi-experiment and survey research), yielding 734 independent samples and 202,468 total participants from 48 countries, studied over a period of 64 years (1950–2014). The study also investigated whether age group and types of county, in particular developed versus developing countries, would act as moderators in terms of intergenerational rise in IQ.

2. Method

2.1. Research instruments

The Raven's Progressive Matrices (RPM), a measure of non-verbal intelligence test that can be used individually or in groups and is widely used in clinical, educational and community settings, was chosen for this study. It is essentially a family of tests that includes three main standardized intelligent test procedures, namely: (a) Coloured Progressive Matrices (CPM); (b) Advanced Progressive Matrices (APM); and (c) Standard Progressive Matrices (SPM).

2.2. Data source and literature search

To investigate the Flynn effect, a cited published article search was conducted using the leading scientific databases ScienceDirect, PubMed, and SpringerLink. The search aim was to obtain all published articles citing the three instruments between 1950 and 2014.

The search terms were “Raven's Progressive Matrices”, “RPM”, “Coloured Progressive Matrices”, “CPM”, “Advanced Progressive Matrices”, “APM”, “Standard Progressive Matrices”, and “SPM”. Over a thousand studies containing targeted data were reviewed and their essential details and characteristics were recorded. This included sample size, year of publication, authors, types of research (experimental versus survey research), country-based participant and types of countries (developed versus developing). The study also recorded the sample mean age, sex ratio (if available) and mean scores and standard deviations for the CPM, APM, and/or SPM (see Table 1). Additionally, occupation was recorded: For the APM dataset, the major occupation is student, accounting for 77.3%, followed by mixed volunteer 6.3%, military 4.1%, and company employee 1.7%; for the CPM, the largest data is again mainly from student (83.9%), followed by mixed volunteer (11.1%), and farmer (0.5%); and for SPM, the major career is student (60.3%), followed by mixed volunteer (7.3%), prisoner (1.1%), and public servant (0.8%).

2.3. Decision rules

Studies were included in the meta-analysis if they reported the mean and/or standard deviation raw scores of the CPM sets A and B, APM set II, and/or SPM sets A to E, and if they employed standard versions of CPM sets A and B, APM set II, and/or SPM sets A to E (excluded were short-form, odd-or even-item versions and modified versions).

Additionally, if studies involved a test–retest method, only mean and/or standard deviation scores for pre-test were recorded and if several articles investigated the same sample or used the same dataset, these statistical parameters (means and standard deviations) were treated as a single data point. Studies were excluded if they investigated clinical research participants (with mental or physical disability) with the

Table 1
Descriptive statistics.

Variables	Mean	SD	Skewness	Kurtosis	Min–max	Number of sample (ns)
Coloured Progressive Matrices						
CPM − mean	22.254	6.226	0.025	−0.932	9.40–35.68	199
CPM − SD	4.622	2.201	5.519	45.786	1.63–24.59	159
Sample size	130.890	253.180	5.160	33.214	6–2255	198
Age	12.627	16.082	3.283	9.574	2.90–80	197
Standard Progressive Matrices						
SPM − mean	39.062	9.657	−0.272	−0.505	11.12–57	369
SPM − SD	7.505	3.304	1.486	7.234	1.06–26.5	256
Age	20.150	14.352	2.237	4.831	5.5–79	344
Sample size	338.16	1389.195	12.206	168.651	6–21432	369
Advanced Progressive Matrices						
APM − mean	23.071	3.112	−0.149	−0.456	14.83–29.83	176
APM − SD	5.141	1.069	−0.388	1.530	0.90–8.51	169
Age	21.961	5.606	2.277	7.053	8.68–46.05	167
Sample size	297.540	675.443	7.396	70.185	8–7335	174
Raven's Progressive Matrices						
RPM − mean	63.965	15.165	−0.181	−0.406	18.53–99.11	734
RPM − SD	6.051	2.871	2.308	11.359	0.90–26.50	575
Age	18.469	13.982	2.349	6.106	2.90–80.00	698
Sample size	271.10	1050.148	14.836	266.139	6–21,432	732

Note: The total number of participants (n); CPN = 25,916; APM: 51,772; SPM: 124,780; and RPM: 202,468.

exception of when they reported data for control groups and these data were appropriate for use. Finally, review articles, letter/personal communication and case report studies were excluded.

The studies were categorize into five age groups (2–5, 6–12, 13–19, 20–39, and >40 years old) using their mean age and also into developed versus developing countries, based on recent International Statistical Institute criteria (ISI) (2014).

2.4. Final sample

The initial search result yielded the following numbers of related-keyword articles respectively for the SpringerLink, Pubmed and ScienceDirect databases: CPM: 602, 133 and 1404; APM: 232, 50 and 744; and SPM: 608, 134 and 1314. Next, guided by the paper title, keywords, and the abstract, duplicated and unrelated articles were removed from the final sample. This stage revealed respectively for CPM, APM and SPM, 351, 522, and 505 articles, out of these 106, 124 and 175 fulfilling the inclusion and exclusion criteria (see Fig. 1).

2.5. Statistical analyses

The study used the Cross-Temporal Meta-Analysis (CTMA) method to analyze the extent to which the RPM mean scores changed over time based on the correlation with the study year. The CTMA method is generally utilized by, for example, by social psychologists (e.g. Twenge, 2000; Twenge & Campbell, 2001, 2008) to analyze trend scores over years. This technique analyzes sample mean data, also weighting the analysis by sample size, so taking into account the fact that larger data-sets should be weighted more highly.

In order to use this method, the following assumptions should be warranted: (a) raw scores must be obtained from widely used tests and with robust psychometrics properties (reliability and validity); and (b) mean scores are weighted by sample size from individual studies to provide unbiased and better estimates of population mean scores. Additionally, to capture generally IQ-score variation, all mean CPM, SPM, and APM raw scores were transformed onto a 0–100 scale, based on standardized data, yielding combined RPM mean scores with uniform scaling (Brouwers et al., 2009).

To quantify the magnitude of change in RPM scores over time, a weighted linear regression was calculated. The year of publication was used as the predictor, with the sample mean RPM score as the outcome measure. Next, to assess the effect of study-level moderating variables on the relationship between predictor and outcome, mean sample age, country types, and their interactions were included in a multiple weighted linear regression model. Also, to take into account within-study variances, a multiple linear regression of all predictors on within-study variances was calculated to determine whether decreasing population variances in mean RPM scores are a cause for IQ gains (Pietschnig et al., 2010).

In addition, mean-centering was used to alleviate the multicollinearity that occurs between lower and higher-order predictor variables for continuous-interaction terms when using multiple regression (Aiken & West, 1991); this also increases the interpretability of regression coefficients (Cohen, Cohen, West, & Aiken, 2003). Finally, to further explore the interaction effects of each of the moderating variables on the Flynn effect, the differences between each pair of correlation coefficients (mean RPM score and year of publication) were analyzed by using the Fisher r-to-z transformation, with the z scores compared using the formula from Cohen and Cohen (1983).

3. Results

Table 1 shows the descriptive statistics, with the mean scores, standard deviation, sample size, and age categorized by CPM (199 samples; n = 25,916), SPM (369 samples; n = 51,772), APM (176 samples; n = 124,780), and RPM (734 samples; n = 202,468). The relatively large range in scores reflects the variability of the large number of studies that were

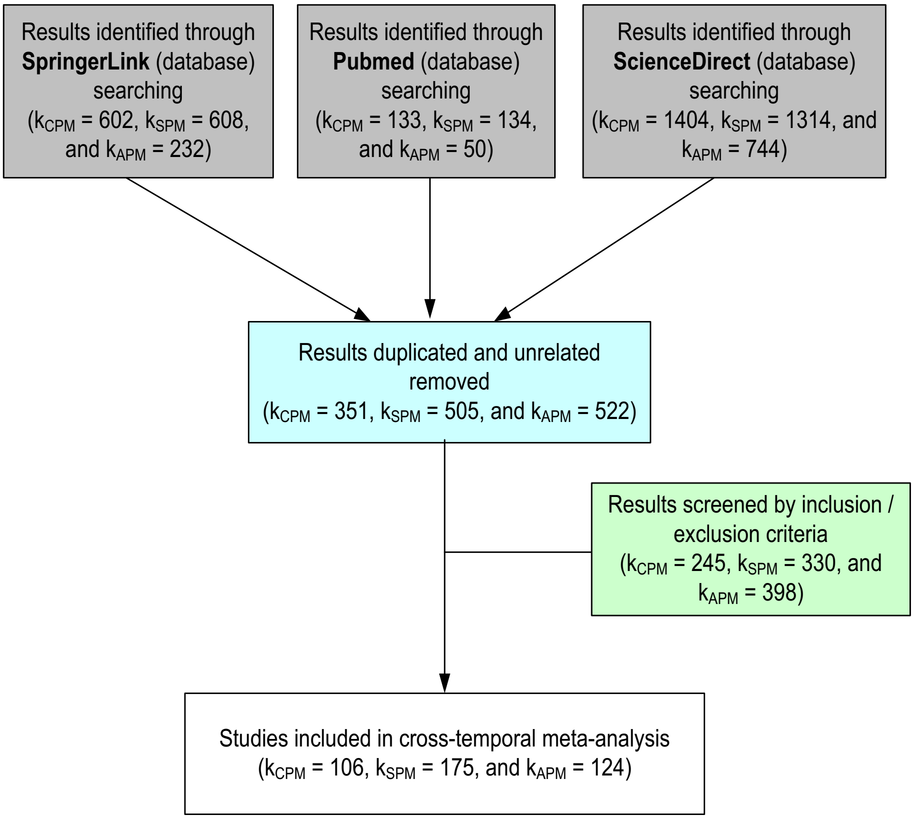

Fig. 1. Flowchart illustrating the process of identifying and screening studies for the CTMA analysis; *k* denoted the number of studies.

employed with their different research paradigms (smaller sample in quasi-experiment) and age groups.

In addition, the current study included the following number of independent samples, according to region: Europe: 285 (271 from developed and 14 from developing countries); North America: 162 (160 from developed and 2 from developing countries); Africa: 127 (all from developing countries); Asia: 117 (79 from developed and 38 from developing countries); South America: 17 (all from developing countries); and Oceania: 22 (all from developed countries) (see Appendix for details).

Table 2 and Fig. 2 clearly demonstrate that the IQ mean scores are significantly correlated with year of publication and with positive correlations for the CPM, SPM, APM, and RPM mean scores when weighed by sample size ($\beta_{CPM} = 0.194$, $p < .01$, $\beta_{SPM} = 0.345$, $p < .01$, $\beta_{APM} = 0.535$, $p < .01$ and $\beta_{RPM} = 0.464$, $p < .01$). Moreover, the effect of mean age of samples indicated that this variable could not predict RPM mean scores, but the main effect of types of countries showed that developed and developing countries had different RPM mean scores. For interaction effects, both age group and types of country moderated the relationship between RPM mean scores and year of publication. Additionally, single and multiple weighted least squares regression analyses for CPM, SPM, APM, and RPM mean scores on within-study variances were conducted. As indicated in Table 3, none of the predictors in the model correlated significantly with the within study variances.

Further analyses, as shown in Table 4 and Fig. 3, indicate that the weighted mean scores of CPM, SPM, APM, and RPM by sample sizes were respectively 20.830, 22.469, 34.228, and 58.442. The moderating effects were present and statistically significant, and yielded different mean RPM scores for each age group and type of countries — 20–39 years old had the highest RPM mean score (65.487), followed by 13–19 years old (59.478), 6–12 years old (59.426), >40 years old (54.392), and 2–5 years old (45.685). Likewise, the magnitude of relationship between RPM mean scores and year of publication was largest in the 13–19 years age group ($r = 0.329$) and smallest, but still statistically significant, in the 30–39 years age group ($r = 0.261$). However, when the differences between each two independent correlation coefficients were tested sequentially, there were no statistical differences among each pair concerning the age groups.

For the types of countries, the Flynn effect size depended on whether the developed or developing countries were considered. Participants in developed countries produced higher scores (30.515 developed *vs* 27.866 developing), but the IQ gain was much larger in developing countries with a correlation coefficient of 0.535, significantly higher than the 0.271 value found for developed countries.

4. Discussion

This study examined the relationship between mean IQ scores and years of publication, spanning 64 years, in which the variables age group and types of country were inputted as potential moderators to explain the given associations. With

Table 2
Single and multiple weighted least squares regression analyses for CPM, SPM, APM, and RPM mean scores.

Variables	ns	B	SE	β	*t*	*p*	R^2	*r*
(*1*) *Coloured Progressive Matrices*								
Single regression	198						0.038	0.194
Intercept		−141.380	58.557		−2.414	0.017		
Year of publication		0.081	0.029	0.194	2.770	0.006		
(*2*) *Advanced Progressive Matrices*								
Single regression	174						0.119	0.345
Intercept		−104.541	26.317		−3.972	<0.001		
Year of publication		0.064	0.013	0.345	4.826	0.006		
(*3*) *Standard Progressive Matrices*								
Single regression	369						0.286	0.535
Intercept		−404.334	36.173		−11.178	<0.001		
Year of publication		0.221	0.018	0.535	12.123	<0.001		
(*4*) *Raven's Progressive Matrices*								
Single regression	731						0.215	0.464
Intercept		−553.220	43.275		−12.784	<0.001		
Year of publication		0.308	0.022	0.464	14.135	<0.001		
Multiple regression	695						0.364	0.603
Intercept		−127.960	70.401		−1.818	0.070		
Mean age of samples (MA)		−0.020	0.039	−0.019	−0.512	0.609		
Type of countries (TC)		−5.948	1.152	−0.204	−5.164	<0.001		
Year of publication (YP)		0.099	0.035	0.127	2.812	0.005		
MA × YP		−0.105	0.038	−0.113	−2.767	0.006		
TC × YP		0.300	0.060	0.210	4.974	<0.001		

Note: Weighted by sample size.

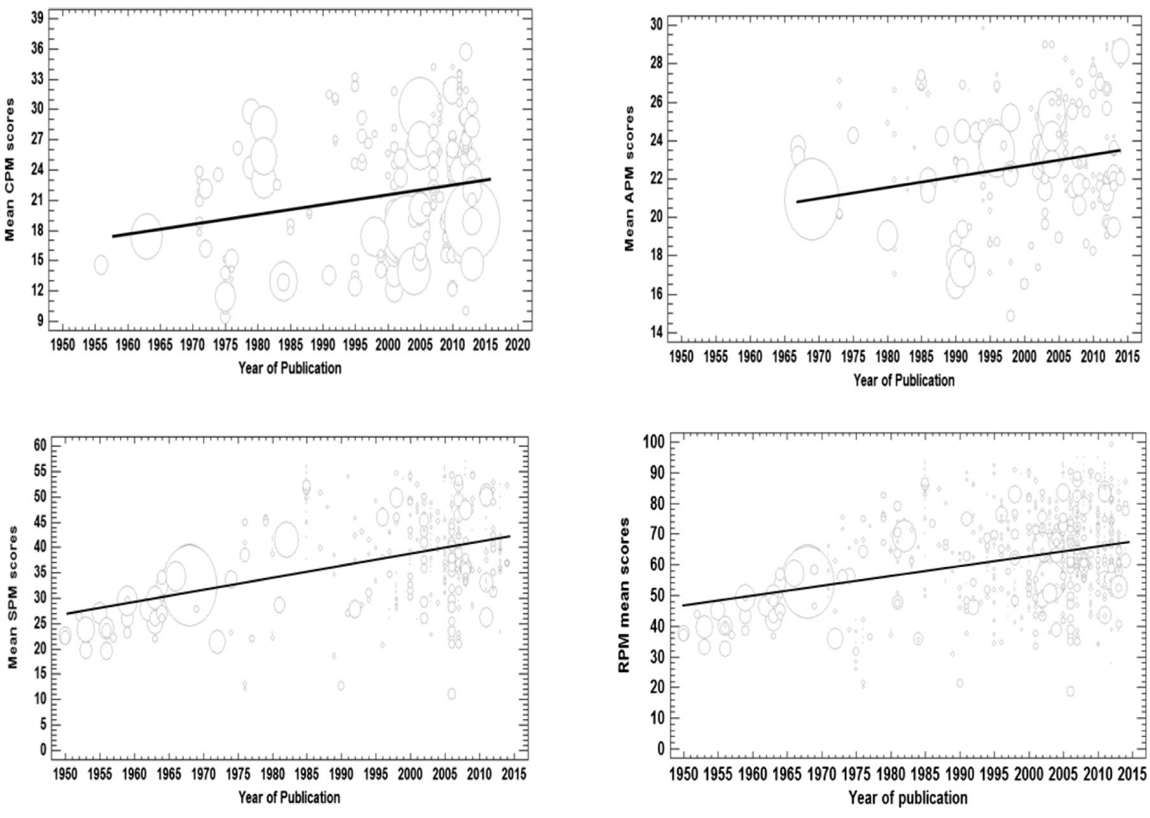

Fig. 2. Regressions between mean RPM scores and years of publication. Upper-left panel: CPM; Upper-right panel: APM; Lower-left panel: SPM; and: Lower-right panel: RPM. The bubble sizes represent visually the relative sample sizes.

Table 3
Single and multiple weighted least squares regression analyses for CPM, SPM, APM, and RPM mean scores on within-study variances.

Variables	ns*	B	SE	β	t	p	R^2	r
(1) *Coloured Progressive Matrices*								
Single regression	159						0.004	0.065
Intercept	(198)	7.814	6.979		1.120	0.265		
Year of publication		−0.003	0.003	−0.065	−0.817	0.415		
(2) *Advanced Progressive Matrices*								
Single regression	160						0.003	0.055
Intercept	(174)	−0.401	3.812		−0.105	0.916		
Year of publication		0.001	0.002	0.055	0.696	0.487		
(3) *Standard Progressive Matrices*								
Single regression	256						0.003	0.057
Intercept	(369)	10.112	8.127		1.244	0.215		
Year of publication		−0.004	0.004	−0.057	−0.915	0.361		
(4) *Raven's Progressive Matrices*								
Single regression	575						0.003	0.056
Intercept	(731)	8.540	4.596		1.858	0.064		
Year of publication		−0.003	0.002	−0.056	−1.335	0.182		
Multiple regression	564						0.021	0.143
Intercept	(695)	3.011	5.472		0.550	0.582		
Mean age of samples		0.003	0.002	0.077	1.702	0.089		
(MA)		0.099	0.061	0.072	1.633	0.103		
Type of countries (TC)		0.000	0.003	−0.007	−0.142	0.887		
Year of publication (YP)		0.001	0.002	0.020	0.450	0.653		
MA × YP		0.011	0.008	0.073	1.439	0.151		
TC × YP								

Note: ns*: number of samples in which SD or variance of sample's means IQ was available for the analysis. Total number of samples are given in parentheses.

the current sizable dataset meta-analysis, the results are favorable in supporting an existence of Flynn effect, but also can be generalized beyond specific types of countries, age group, and sample size or research methods, an area of concern raised in the critique by Rodgers (1999).

Overall, it was clear that the Flynn effect could be observed on RPM tests. In this respect, our results were consistent with many previous meta-analysis findings (Brouwers et al., 2009; Pietschnig et al., 2010). However, the results have several and important points of interest regarding the overall effect.

Table 4
RPM mean scores and correlations with year of publication for RPM types, age groups and types of countries.

Target variables (ns)	Mean scores (weighted by sample size)	Correlation coefficient between mean IQ test scores and year of publication (weighted by sample size)
CPM (198)	20.830	0.194*
APM (174)	22.469	0.345**
SPM (369)	34.228	0.535**
RPM (731)	58.442	0.464**
[1]Age group (2–5 years old) (45)	45.685	0.296*
[2]Age group (6–12 years old) (243)	59.423	0.325**
[3]Age group (13–19 years old) (136)	59.479	0.329**
[4]Age group (20–39 years old) (222)	65.487	0.261**
[5]Age group (>40 years old) (49)	54.392	0.308*
Developed countries (532)	30.515	0.271**
Developing countries (198)	27.866	0.535**

Note: Test of difference between correlations for age groups and also developed versus developing countries.
$r_{1\text{ vs }2}$, z = −0.192, p = 0.845 (two-tailed).
$r_{1\text{ vs }3}$, z = −0.207, p = 0.836 (two-tailed).
$r_{1\text{ vs }4}$, z = 0.225, p = 0.822 (two-tailed).
$r_{1\text{ vs }5}$, z = −0.062, p = 0.951 (two-tailed).
$r_{2\text{ vs }3}$, z = −0.041, p = 0.967 (two-tailed).
$r_{2\text{ vs }4}$, z = 0.75, p = 0.453 (two-tailed).
$r_{2\text{ vs }5}$, z = 0.117, p = 0.907 (two-tailed).
$r_{3\text{ vs }4}$, z = 0.678, p = 0.498 (two-tailed).
$r_{3\text{ vs }5}$, z = 0.137, p = 0.891 (two-tailed).
$r_{4\text{ vs }5}$, z = −0.315, p = 0.752 (two-tailed).
$r_{\text{developed vs developing}}$, z = −3.810, p < 0.01 (two-tailed).
* p < .05.
** p < .01.

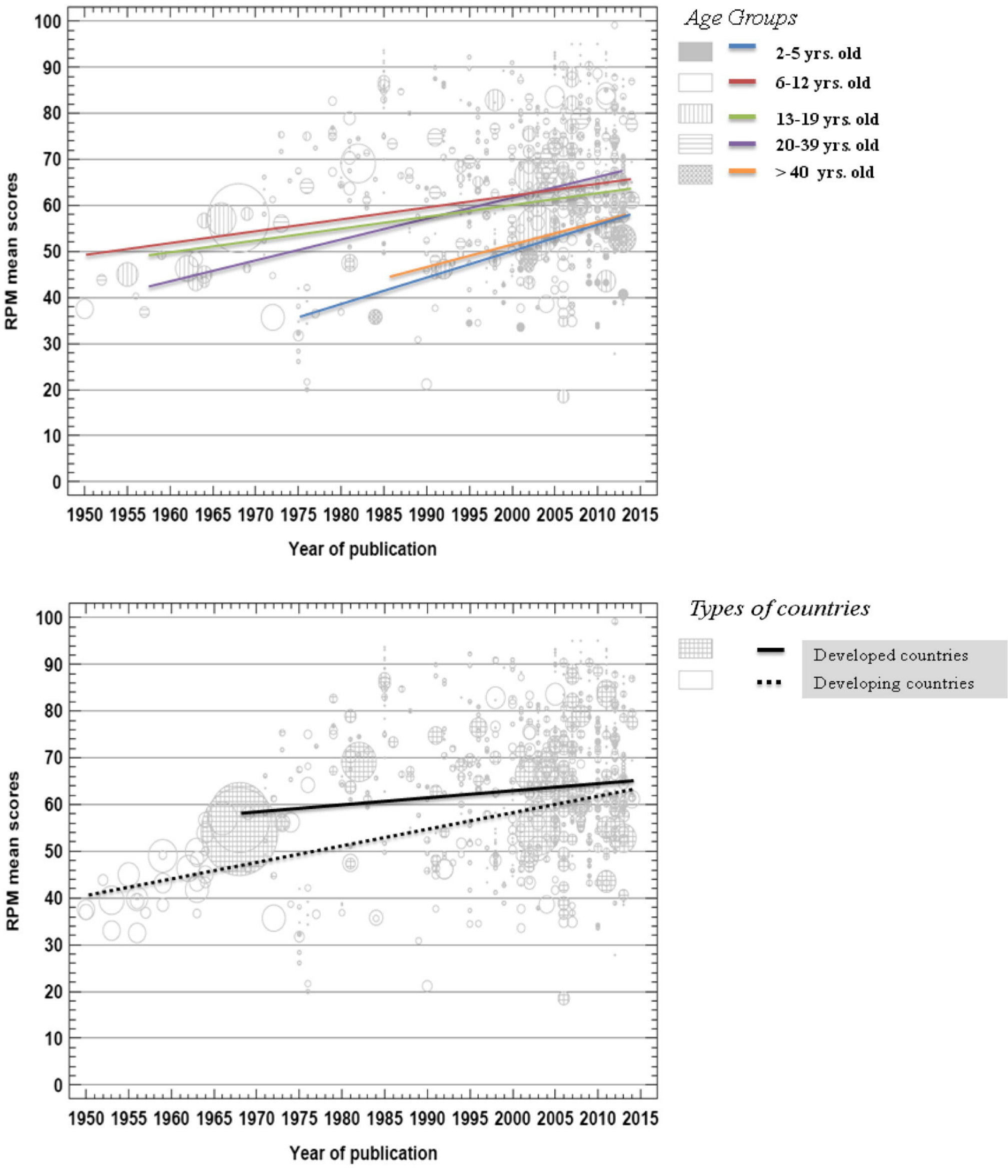

Fig. 3. Weighted regression of the relationship between mean RPM scores and year of publication. Upper panel: weighted by sample sizes and divided by age groups (2–5 years old, 6–12 years old, 13–19 years old, 20–39 years old, 20–39 years old and more than 40 years old). Lower panel: weighted by sample sizes and divided by type of country (developed and developing countries). The bubble sizes represent visually the relative sample sizes.

Firstly, though the Flynn effect was present in all RPM tests, the SPM distinctively outperformed other tests ($SPM_{r\,=\,0.535}$, $APM_{r\,=\,0.345}$, and $CPM_{r\,=\,0.194}$). This may reflect the psychometric properties of SPM, being designed to be suitable for the general population, the APM in contrast having been designed for adolescents and adults of above-average intelligence, with participants achieving higher scores on the SPM than on the APM. For CPM, we suspect that sample heterogeneity may somewhat confound interpretation of the results. For example, when older adult participants were removed from the dataset, the correlation coefficient between years of publication and CPM mean scores increased from 0.194 to 0.221. In addition, there is the fact that the SPM is considered the most widely used among RPM tests, with the largest meta-analysis sample; increased exposure to this particular test might have caused more IQ gain.

Secondly, the current study investigated whether pervasiveness of IQ rise is dependent on age. Although the tests of statistical differences between pairs of correlation coefficient did not reveal any significant effects, this interpretation may require caution, especially in case of significantly interaction effects between continuous variables on multiple regression analysis. There are also various problems that could arise (Aiken & West, 1991), particularly when a main effect is not statistically significant, in which the main effect and interaction effect can be confounded.

On the other hand, it is worth noting that the Flynn effect was found for every age group. High school students showed the effect and IQ gains were also robust for the adult samples. Flynn (2013) has recently suggested that the Flynn effect should be found in all age ranges, including the very young, and his data (Flynn, 2012) also showed that IQ gains in 2008 were large for all age groups, with the exception of those aged 14.5–15.5 years.

Thirdly, the larger Flynn effect in developing countries suggests that IQ is catching up with the developed world, the gap closing to 2.649 points. There are multiple possible reasons proposed for this phenomenon, including for example, improved education, increased exposure to testing, improved medical care and nutrition, heterosis, decreased family size, modernized child rearing and educational practices, artificial light and more complex cognitively simulating environments (Williams, 2013 for review). It should also be noted, however, that global conclusions about Flynn effect regarding developing countries have to be regarded cautiously because of their diverse characteristics. These include whether the social or geopolitical characteristics of the countries are comparable in their potential to produce change and also whether a developing country will transform into a developed country over time, further complicating interpretation. In addition, it is important to consider the potential contribution of ceiling effects on SPM measurement (Mackintosh & Bennett, 2005), which might limit the ability of the IQ test to detect higher functioning and so give a false impression of low rate of increase in the higher scoring developed country populations. Indeed, a higher proportion of individuals in samples from developed countries had maximum scores on the SPM.

In conclusion, the CTMA method was used in an attempt to produce a less biased estimate of the Flynn effect. With this method, a full range of age groups, countries and research methodologies (quasi-experimental design procedure, cohort study, and survey research) were all systematically evaluated, applied to the RPM tests. Over the course of decades, the gain in IQ score is generally robust and was found to be predominantly present in developing countries and more so for the SPM. The Flynn effect is strong enough to be showed in even small and non-representative samples as well as in the very young and older adults.

Contributors

Peera Wongupparaj, Veena Kumari and Robin G. Morris designed the study. Peera Wongupparaj undertook the statistical analysis and prepared the first draft. All authors contributed to and approved the final manuscript.

Acknowledgments

This study was supported by Thai Government PhD scholarship (Grant no. CS4198) and funds were from Biomedical Research Centre for Mental Health at the Institute of Psychiatry, Psychology, and Neuroscience, King's College London and South London and Maudsley NHS Foundation Trust for some of VK's time. We are grateful to Professor James Flynn and an anonymous reviewer for reviewing the paper and providing suggestions for improvement.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.intell.2014.11.008.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park: Sage.

Batterjee, A. A., Khaleefa, O., Ali, K., & Lynn, R. (2013). An increase of intelligence in Saudi Arabia, 1977–2010. *Intelligence*, *41*, 91–93.

Brouwers, S. A., Van de Vijer, F. J. R., & Van Hemert, D. A. (2009). Variation in Raven's Progressive Matrices scores across time and space. *Learning and Individual Differences*, *19*, 330–338.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Cotton, S. M., Kiely, P. M., Crewther, D. P., Thomson, B., Laycock, R., & Crewther, S. G. (2005). A normative and reliability study for the Raven's Colored Progressive Matrices for primary school aged children in Australia. *Personality and Individual Differences*, *39*, 647–660.

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

Flynn, J. R. (2012). *Are we getting smarter: Rising IQ in the twenty-first century*. Cambridge University Press.

Flynn, J. R. (2013). The "Flynn Effect" and Flynn's paradox. *Intelligence*, *41*(6), 851–857.

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

Liu, J., & Lynn, R. (2013). An increase of intelligence in China 1986–2012. *Intelligence*, *41*, 479–481.

Liu, J., Yang, H., Li, L., Chen, T., & Lynn, R. (2012). An increase of intelligence measured by the WPPSI in China, 1984–2006. *Intelligence*, *40*, 139–144.

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

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

Lynn, R., & Harvey, J. (2008). The decline of the world's IQ. *Intelligence*, *36*, 112–120.

Mackintosh, N. J., & Bennett, E. S. (2005). What do Raven's Matrices measure? An analysis in terms of sex differences. *Intelligence*, *33*, 663–674.

Pietschnig, J., Voracek, M., & Formann, A. K. (2010). Pervasiveness of IQ rise: A cross-temporal meta-analysis. *PLoS one*, *5*(12), 1–6.

Rodgers, J. L. (1999). A critique of the Flynn Effect: Massive IQ gains, methodological artifacts, or both? *Intelligence*, *26*(4), 337–356.

Rushton, J. P., Skuy, M., & Bons, T. A. (2004). Construct validity of Raven's Advanced Progressive Matrices for African and non-African engineering students in South Africa. *International Journal of Selection and Assessment*, *12*, 220–229.

Rushton, J. P., Skuy, M., & Fridjhon, P. (2003). Performance on Raven's Advanced Progressive Matrices by African, East Indian, and White engineering students in South Africa. *Intelligence*, *31*(2), 123–137.

Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004a). 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.

Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004b). 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. T., Murphy, R., & van Eeden, R. (2011). The Flynn effect in South Africa. *Intelligence*, *39*, 456–467.

te Nijenhuis, J., & van der Flier, H. (2013). Is the Flynn effect on g?: A meta-analysis. *Intelligence*, *41*, 802–807.

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

The International Statistical Institute (2014). Developing countries. Retrieved June 10, 2014, from http://www.isi-web.org/component/content/article/5-root/root/81-developing.

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

Twenge, J. M. (2000). The age of anxiety? Birth cohort change in anxiety and neuroticism, 1952–1993. *Journal of Personality and Social Psychology*, *79*, 1007–1021.

Twenge, J. M., & Campbell, W. K. (2001). Age and birth cohort differences in self-esteem: A cross-temporal meta-analysis. *Personality and Social Psychology Review*, *5*, 321–344.

Twenge, J. M., & Campbell, S. M. (2008). Generational differences in psychological traits and their impact on the workplace. *Journal of Managerial Psychology*, *23*(8), 862–877.
Van de Vijver, F. J. R. (1997). Meta-analysis of cross-cultural comparison of cognitive test performance. *Journal of Cross-Cultural Psychology*, *28*, 678–709.
Williams, R. L. (2013). Overview of the Flynn effect. *Intelligence*, *41*, 753–764.
Woodley, M. A., & Meisenberg, G. (2013). A Jensen effect on dysgenic fertility: An analysis involving the National Longitudinal Survey of Youth. *Personality and Individual Differences*, *55*, 279–282.