Regional and Sex Differences on the Standard Progressive Matrices in the Kingdom of Saudi Arabia

Ibraheem Abdulaziz Abdullah Almuaqel
King Saud University, Department of Special Education, College of Education, Riyadh, Saudi Arabia
E-mail: muaqel@ksu.edu.sa

Edward Dutton
Asbiro University, Łódź, Poland
E-mail: ecdutton@hotmail.com

Salaheldin Farah Attallah Bakhiet*
King Saud University, Department of Special Education, College of Education, Riyadh, Saudi Arabia
*Corresponding author; E-mail: bakhiet@ksu.edu.sa

David Becker
Department of Psychology, Chemnitz University of Technology, Germany
E-mail: david.becker@s2009.tu-chemnitz.de

Zeinab Abdel Rahman Elhassn Hagge Ahmed
Department of Kindergarten, College of Science and Humanities, Prince Sattam bin Abdulaziz University, Al-Aflaj, Saudi Arabia
E-mail: Z.haggeahmed@psau.edu.sa

This study examines the interaction between gender and region in Saudi Arabia on performance in the Standard Progressive Matrices (SPM), a test of non-verbal reasoning. The results of two administrations of the SPM in Saudi Arabia are compared for children between the ages of 8 and 12 years. The first sample included 1,273 school children in this age range, tested in the Makkah region in 2010. The second included 2,385 school children in Riyadh in 2016. There was no statistically significant IQ difference between the two samples. However, females scored significantly higher in Riyadh than in Makkah while males scored significantly higher in Makkah. These differences became more pronounced with age.

Keywords: Saudi Arabia, Intelligence, Regional differences, Sex differences

This study explores two related issues in the Kingdom of Saudi Arabia. These are sex differences in intelligence, and regional differences in intelligence. There has been much recent debate over whether sex differences in IQ actually exist among adults, but it is agreed that females often have a small IQ advantage in early adolescence, perhaps because they enter puberty earlier leading to earlier cognitive development (Flynn, 2012, 2017; Lynn, 2017a,b; Lynn & Irwing, 2004; van der Linden et al., 2017). However, these findings are based on studies in developed countries. Here we wish to contribute to the understanding of sex differences in cognitive test scores by exploring the issue in an Arabic country. Bakhiet et al. (2018) have already shown that the average IQ of schoolgirls in Arab countries is consistently higher than that of boys, certainly where primary education is compulsory. This pattern has been found in Yemen (Bakhiet et al., 2015), the United Arab Emirates (Khaleefa & Lynn, 2008), Oman (Khaleefa et al., 2012), and Sudan (Bakhiet & Lynn, 2014). Our study contributes to this field of research by providing a further data point for future meta-analysts concerned with understanding sex differences in intelligence in the Arab world and their possible causes.

The present study will contribute to the understanding of sex differences in child IQ scores in the Arabic world, by investigating whether they can be found when comparing two large metropolitan regions in the Kingdom of Saudi Arabia: its centrally located capital Riyadh, and the coastal region of Makkah. The latter is a wealthy and cosmopolitan region which includes Mecca and the port city of Jeddah. These are two very different regions within the country, meaning that comparing them is potentially informative. We will compare two large samples of school children aged 8-12 years on the Standard Progressive Matrices. This will also permit us to explore regional differences in IQ, in addition to sex differences.

Many studies have already demonstrated that there exist regional differences in IQ within many countries. The usual finding has been that regions with higher average IQ tend to score higher on assorted measures of socio-economic development. This has been demonstrated in the regions of the UK (e.g. Carl, 2015; Lynn, 1979), France (Lynn, 1980), Italy (e.g. Lynn, 2010), Spain (Lynn, 2012), Portugal (Almeida et al., 2011), Germany (Roivainen, 2012), Finland (Dutton & Lynn, 2014), China (Lynn & Cheng, 2013), Japan (Kura, 2013), the

USA (e.g. Pesta et al., 2010), Turkey (Lynn et al., 2015), Brazil (Fuerst & Kirkegaard, 2016), Mexico (Fuerst & Kirkegaard, 2016), Sudan (Bakhiet & Lynn, 2014), and Russia (Grigoriev et al., 2016). This study will provide a further data point for future meta-analysts.

In addition, the results of both dimensions of this study will allow us to present possible hypotheses for explaining our findings, which it may be possible to test in future studies. One tentative hypothesis is that we might expect people to score higher in Riyadh because it is the Saudi capital and we would expect more intelligent people to be attracted to the capital region for work reasons, although in this case the "provincial" region of Makkah is more "cosmopolitan" than the capital city of Riyadh in many respects. We would also expect males to attain a slight IQ advantage over females towards their late teens as has been noted in Western countries (Lynn, 2017a), though this is subject to much debate (Flynn, 2017). However, in the 8-12 years age range, we predict a small female advantage in accordance with Richard Lynn's developmental theory of sex differences (2017a). We also expect that the gender difference is affected by cultural conditions and gender roles in these two locations, although the difficulty predicting the net direction of such effects necessitates a data-driven approach.

Method

Sample 1

The Makkah sample consisted of 3,209 Saudi students aged between 8 and 18 years from the Makkah region, of which 1,613 (50.2%) were males and 1,596 (49.8%) were females. The mean age was 12.3 and the standard deviation was 4.6. The total sample was divided into 11 age groups aged 8 to 18 years. The sample was stratified, selected from state schools in the region's three main cities: Mecca, Taif, and Jeddah. The data collection process took place between 24 April and 23 June 2010. An analysis of these data has been published, in Arabic, in Batterjee (2014).

Sample 2

The second sample consisted of 2,945 school students attending state schools in Riyadh, the capital of Saudi Arabia. Due to missing data and coding errors, adequate test results could only be determined for 2,385 of these. This reduced sample was composed of 1,263 (53%) boys and 1,122 (47%) girls aged 6 to 12 years. The mean age was 10.36 and the standard deviation was 1.33. In the following we compare the two administrations for those aged between 8 and 12 years. The sample was stratified, selected from schools in the north, south, east, west, and center of Riyadh, to represent all economic and social levels in the city. The data were collected by some of the present authors between February and June 2016.

Data collection

In both studies, two teams of graduate research assistants were trained: a male team for male schools, and a female team for female schools. Both teams were trained in administering the test, correcting it, monitoring grades using special forms and files, and entering the information into the relevant computer program. Before test administration, approval was obtained from the Education Administration in each region. The idea and objectives of the research were explained to students, and those not wanting to take the exam were allowed to leave class. The region, whether Makkah or Riyadh, was divided into 5 units (North, South, East, West, and Central), so that this division is comprehensive for all parts of the region. The samples were recruited in governmental and private schools by the random stratified method, and the test was applied according to the instructions in the test manual. Comprehensive explanation was provided to teachers in the schools about the research and testing and how to apply it. The teachers participated in the process of testing in the classroom, and researchers assisted in distributing questionnaires, answer sheets and pens, as the tests were administered collectively in the students' classrooms.

After the data collection process, incomplete answer sheets were excluded and the research assistants entered the study data into the computer program. Then the statistician analyzed the data and extracted the results and included them in the study report. Data analysis was conducted using SPSS.

Instrument

The Standard Progressive Matrices (SPM) was designed by Raven (1941). The test contains five parts (A, B, C, D, and E), each consisting of 12 matrices, where each matrix is composed of drawings or designs. The difficulty of the puzzles within each group increases gradually until the end of the test. The test is suitable for ages from 6 to 60 years and beyond.

Each item is a rectangle with drawings or shapes that are systematically different from each other going from left to right and top to bottom, according to criteria that the examinee has to recognize. One drawing is missing, and the examinee must fill in the deleted part from six or eight options given at the bottom of the rectangle. In parts A and B the number of options is six, and in the others the number of options is eight. Part A begins with easy tasks requiring fill-in of the cut-out piece of a geometric pattern. Part B also progresses from easy to difficult, and this group, as a whole, is more difficult than part A. Like part A, it tests visual ability but focusing on the subject's ability to identify symmetry between the forms.

The SPM test scores have high internal consistency, with Cronbach's alpha typically ranging between .80 and .90 in different samples from non-western countries (e.g., Bakhiet, 2008; Batterjee & Ashria, 2015; Humble et al., 2016; Hur & Lynn, 2013; Husain et al., 2019; Owen, 1992). Concurrent validity with verbal and performance intelligence tests ranges between .40 and .75. The test has been applied in many countries and on large groups of examinees between the ages of 6 and 65 years (Raven, Raven & Court, 1985, 1998), and the test-retest reliability ranged between .80 and .90.

Results

Information about SPM raw scores is given in Table 1 for Riyadh and Makkah.

Table 1. *SPM raw scores in Riyadh (2016) and Makkah (2010).*

Age	Sex	Riyadh		Makkah		Difference
		N	*M ± SD*	*N*	*M ± SD*	
8	Males	141	15.35 ± 8.92	81	14.93 ± 7.66	0.42
	Females	131	17.68 ± 8.36	142	16.78 ± 8.01	0.90
	Both	272	16.47 ± 8.73	223	16.11 ± 7.94	0.36
9	Males	254	17.38 ± 8.31	74	19.74 ± 9.54	-2.37
	Females	148	25.33 ± 9.21	123	23.54 ± 9.80	1.79
	Both	401	20.32 ± 9.46	197	22.11 ± 9.85	-1.79
10	Males	264	19.10 ± 10.56	102	24.20 ± 9.98	-5.10
	Females	280	28.30 ± 7.76	134	28.40 ± 8.80	-0.10
	Both	544	23.83 ± 10.15	236	26.58 ± 9.54	-2.75
11	Males	279	21.69 ± 11.49	172	28.47 ± 8.83	-6.78
	Females	257	32.20 ± 9.25	129	29.71 ± 8.55	2.49
	Both	536	26.73 ± 11.72	301	29.00 ± 8.72	-2.27
12	Males	326	20.61 ± 10.44	169	30.35 ± 7.71	-9.74
	Females	306	32.27 ± 9.45	147	30.53 ± 8.74	1.75
	Both	632	26.26 ± 11.55	316	30.43 ± 8.19	-4.17
All	Males	1263	19.30 ± 10.37	598	24.87 ± 8.73	-5.57
	Females	1122	28.65 ± 10.02	675	25.19 ± 8.56	3.46
	Both	2385	23.70 ± 11.22	1273	24.79 ± 8.44	-1.09

Discussion: Table 1

An age-related increase can be seen in both regions, with an average annual gain in Riyadh of 2.45 and in Makkah of 3.58 raw score points. Raven (2000, B1) reported a stronger annual increase of 4.00 at the 50th percentile of the British norm sample of 1979. However, the stepwise gains in Britain were +7 from ages 8 to 9, +6 from age 9 to 10, +2 for age 10 to 11, and +1 for age 11 to 12. In Saudi Arabia stepwise gains are similar to those in the British norms within the ages of 8 to 11, with an annual increase of 3.42 in Riyadh and 4.30 in Makkah.

The mean raw score of 25.29 measured in Riyadh is around the 10th percentile of the 1979 British norms and therefore equivalent to a British 1979 IQ of 81. The mean raw score of 24.79 measured in Makkah is between the 12th and 13th percentile of the 1979 British norms and therefore equivalent to an IQ of 82-83. The term Flynn Effect is used to describe a secular increase in IQ scores between cohorts. The meta-analysis by Pietschnig and Voracek (2015, S2) reported a Flynn Effect in fluid intelligence in Britain between 1979 and 2010/16 of around 6.1 IQ points (annual gain = 0.21; time-span: 1979-2008), but Flynn (2009) mentioned a much higher Flynn Effect of 14 for 5-15 year olds.

Table 2. *Standardized difference* d *in intelligence between Riyadh (2016) and Makkah (2010); p values from two-tailed* t*-test.*

Age	Sex	Riyadh (2016) vs. Makkah (2010)			
		d	*t*	*p*	*F*
8	Males	0.06	0.353	.724	1.356
	Females	0.11	0.908	.365	1.088
	Both	0.04	0.476	.634	1.209
9	Males	-0.27	-2.074	.039	1.319
	Females	0.19	1.548	.123	1.132
	Both	-0.19	-2.149	.032	1.083
10	Males	-0.49	-4.208	<.001	1.119
	Females	-0.01	-0.117	.907	1.287
	Both	-0.28	-3.494	.001	1.167
11	Males	-0.64	-6.626	<.001	1.694***
	Females	0.28	2.560	.011	1.171
	Both	-0.21	-2.935	.003	1.807***
12	Males	-1.01	-10.713	<.001	1.832***
	Females	0.19	1.884	.060	1.169
	Both	-0.40	-5.741	<.001	1.989***
All	Males	-0.56	-11.373	<.001	1.410***
	Females	0.36	7.472	<.001	1.369***
	Both	-0.11	-3.050	.002	1.767***

Notes: Negative Cohen's *d* represent higher scores for Makkah; critical values for Cohen's *d*: ±.41 = small, ±1.15 = medium, ±2.70 = strong; *t*-test unpaired and two-tailed; critical values for *F*-test for differences in *SD*: * $p \leq .050$; ** $p \leq .010$; *** $p \leq .001$.

Discussion: Table 2

For differences in mean scores, we used *t*-tests and Cohen's *d*. Due to the lack of other comparative samples, we used $p < .001$ as the minimum level of significance and *d*-conventions for interpretation of effect sizes from Ferguson (2009), which are stricter than those from Cohen (1988, p. 179-213). We found statistically significant differences in raw scores with small effect sizes between Riyadh and Makkah for males in the full sample and in ages 10 to 12, with higher scores in Makkah (Table 2, Fig. 1). In contrast, girls from Riyadh outperform girls from Makkah, but without statistical significance except in the full sample, were the *t*-test gives a $p < .001$. Not only sex differences but also individual differences, measured by the standard deviations, are higher in Riyadh at the older ages (see Table 1). The nature of the sex differences can be seen in more detail in Table 3 and Figure 1.

Table 3. *Sex differences in SPM raw scores in Riyadh (2016) and Makkah (2010). F test is for comparison of standard deviations.*

Age	Riyadh (2016)			Makkah (2010)		
	d	*p*	*F*	*d*	*p*	*F*
8	-0.27	.027	1.240	-0.23	.093	1.093
9	-0.92*	<.001	1.230	-0.39	.008	1.055
10	-1.00*	<.001	1.853***	-0.45*	.001	1.286
11	-1.00*	<.001	1.543***	-0.14	.223	1.067
12	-1.17**	<.001	1.220*	-0.02	.846	1.285
All	-0.92*	<.001	1.071	-0.04	.510	1.040

Notes: Negative Cohen's *d* represent higher scores for females; critical values for Cohen's *d*: ±.41* = small, ±1.15** = medium, ±2.70*** = strong; *p* value is for *t*-test unpaired and two-tailed; critical values for *F*-test for differences in *SD*: * $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$.

Discussion: Table 3

Table 3 shows substantial effect sizes in Riyadh, with higher mean scores for females at all ages. F-test shows no significant male-favoring sex-differences in standard deviations for the full sample but a high significance in the case of 10- and 11-year-olds and a weak significance in the case of 12-year-olds. In Makkah, effect sizes of sex on mean scores also favor females, but do not exceed the critical value for a minimum effect except in the age group of 10-year-olds. For the whole sample, sex differences in mean scores are negligible. Similarly, F-tests show no significant sex differences in standard deviations in Makkah.

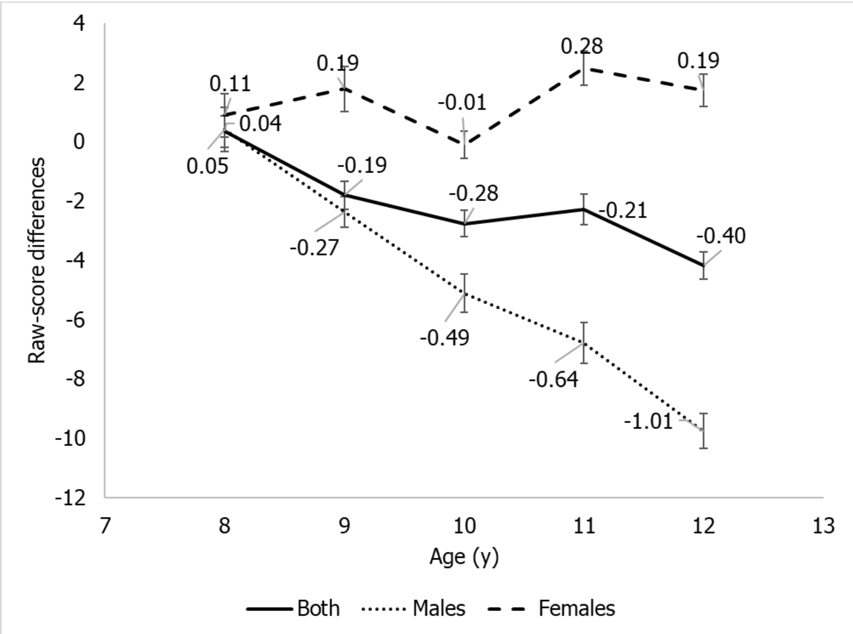

Figure 1. *Comparison of raw scores between Makkah (2010) and Riyadh (2016). Positive differences represent higher scores for Riyadh; error bars represent S.E. for differences between independent means; Cohen's d as numbers at the marks.*

General Discussion

Overall, we find no significant IQ difference between Riyadh and Makkah. This may be because both are relatively prosperous regions of the country, attracting relatively intelligent people from the rest of the country in each generation, though this hypothesis needs to be tested. However, this lack of significant differences in explained by girls, who performed at similar level in Riyadh and Makkah. In contrast, boys from Makkah outperform boys from Riyadh, and the more so, the older they get. Another point of interest is that the female IQ advantage in this age range is replicated but is far greater in Riyadh than in Makkah. This female advantage is unusually large in Riyadh also when compared to other countries, especially among those aged 11-12. For example, administration of the SPM in Estonia found a female advantage of 0.54 standard deviations at age 12 (Lynn et al., 2004). In summary, the speed of cognitive development of boys is much lower in Riyadh compared to Makkah, but girls show no differences between the regions, neither in ability level nor in cognitive development.

Why should there be a female advantage and why should it be so much more pronounced in Riyadh? The female advantage is not in itself surprising in any country with compulsory education. Females mature faster than males physically as evidenced by an earlier age at puberty. It has been hypothesized that brain maturation follows the same schedule as maturation of other organs during adolescence, with the consequence that for a few years during development, females have higher average IQs than males (e.g., Lynn, 2017a). This is approximately the age range with which we are dealing in these samples. There remains considerable debate over whether adult females have lower IQ than males as we have already seen; but there is no dispute over data indicating a female advantage, of varying degrees, among children or young teenagers. It is peculiar, however, that the sex difference in Riyadh is so big. It is negligible in many Western studies of children of this age (Flynn, 2012, Appendix IV).

Serious bias in selecting the Riyadh sample is unlikely. Boys and girls attend different schools, but when choosing the samples, we were keen to take a school for boys and a school for girls from the same area in order to ensure that they represent the same cultural, economic and social level, and the same professional status of families. All schools were public schools, administered by the Ministry of Education. That is, the selection of the sample was balanced so that it represents the socio-economic levels in the city of Riyadh, without bias to the richer groups or others.

Also, girls' and boys' schools teach the same curricula. Therefore, the more likely explanation is the personal interest of girls in school work. Partly as a result of cultural constraints on social and outdoor activities of girls, they spend more time than boys in their homes and therefore spend more time studying, which makes their cognitive abilities slightly higher than those of boys. This may be different at more advanced ages and at higher educational levels in secondary school and university.

Another point of interest is the steeper increase in raw scores from the age of 11 to 12 in Saudi Arabia, compared to the increase in the British standardization sample. This is remarkable against the background of an otherwise smaller overall increase from age 6 to 12. It may be explicable as the result of a delayed timing of cognitive development in Saudi Arabian children. This is consistent with evidence that the average age of menarche is later in Saudi Arabia, 13.08 years on average (Al Alwan et al., 2015), than the UK average of 12.7 years (NHS, 2011). However, those British norms were estimated in the late 1970s and therefore allow only limited comparison.

As the SPM continues to be used widely in cross-national studies (see studies collected by Lynn & Vanhanen, 2002, 2012), a new standardization on a representative sample in the UK would be appropriate. The SPM+ standardization from 2007 (Raven, 2008) could alternatively be used, but a validated conversion table is missing. It would be a meaningful task to recruit a sample of at least 300 subjects and apply randomly the SPM to half of them and the SPM+ to the other half, or administer both tests to all of them in counterbalanced sequence, to create such a table.

Conflicts of Interests: The authors declare that they have no conflict of interest.

Acknowledgements: This research was funded by the Deanship of Scientific Research at Prince Sattam bin Abdulaziz University.

References

Al Alwan, I., Ibrahim, A., Badri, M., Al Dubayee, M. & Bin-Abbas, B. (2015). Decline in menarcheal age among Saudi girls. *Saudi Medical Journal* 36: 1324-1328.

Almeida, L., Lemos, G. & Lynn, R. (2011). Regional differences in intelligence and per capita incomes in Portugal. *Mankind Quarterly* 53: 213-221.

Bakhiet, S.F.A. (2008). Application of Progressive Matrices in the Sudan. *Arab Psynet* 19: 223-225.

Bakhiet, S. & Lynn, R. (2014). Regional differences in intelligence in Sudan. *Intelligence* 50: 150-152.

Bakhiet, S., Al-Khadher, M. & Lynn, R. (2015). Study of means and sex differences on Raven's Standard Progressive Matrices Plus in Yemen. *Mankind Quarterly* 55: 268-277.

Bakhiet, S., Dutton, E., Ashaer, K., Essa, Y., Blahmar, T., Hakami, S. & Madison, G. (2018). Understanding the Simber Effect: Why is the age-dependent increase in children's cognitive ability smaller in Arab countries than in Britain? *Personality and Individual Differences* 122: 38-42.

Batterjee, A. (2014). The impact of educational quality and some demographic variables on the intelligence of general education students in the Makkah region. *Message of Education and Psychology* 46: 159-187.

Batterjee, A.A. & Ashria, I. (2015). Intelligence and education: The Sudan case. *Mankind Quarterly* 56: 135-166.

Carl, N. (2015). IQ and socioeconomic development across regions of the UK. *Journal of Biosocial Science* 48: 406-417.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd edition. Lawrence Erlbaum.

Dutton, E. & Lynn, R. (2014). Regional differences in intelligence and their social and economic correlates in Finland. *Mankind Quarterly* 54: 447-456.

Ferguson, C.J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice* 40: 532-538.

Flynn, J. (2009). Requiem for nutrition as the cause of IQ gains: Raven's gains in Britain 1938–2008. *Economics & Human Biology* 7(1): 18-27.

Flynn, J. (2012). *Are We Getting Smarter? Rising IQ in the Twenty-First Century.* Cambridge: Cambridge Univ. Press.

Flynn, J. (2017). Male and female balance sheet. *Mankind Quarterly* 58: 43-68.

Fuerst, J. & Kirkegaard, E. (2016). Admixture in the Americas and national differences. *Mankind Quarterly* 56: 256-374.

Grigoriev, A., Lapteva, E. & Lynn, R. (2016). Regional differences in intelligence, infant mortality, stature and fertility in European Russia in the late nineteenth century. *Intelligence* 55: 34-37.

Humble, S., Dixon, P. & Schagen, I. (2016). Assessing intellectual potential in Tanzanian children in poor areas of Dar es Salaam. *Assessment in Education: Principles, Policy & Practice* 25(4): 351-354. doi:10.1080/0969594X.2016.1194257.

Hur, Y.-M. & Lynn, R. (2013). Twin-singleton differences in cognitive abilities in a sample of Africans in Nigeria. *Twin Research and Human Genetics* 16: 1-8.

Husain, N.I.A.E., Meisenberg, G., Becker, D., Bakhiet, S.F., Essa, Y. Ah. S., Lynn, R. & Al Julayghim, F.M.H. (2019). Intelligence, family income and parental education in the Sudan. *Intelligence* 77: 101402. doi:10.1016/j.intell.2019.101402

Khaleefa, O. & Lynn, R. (2008). A study of the intelligence of the United Arab Emirates. *Mankind Quarterly* 49: 58-64.

Khaleefa, O., Al-Kudri, K. & Lynn, R. (2012). Intelligence in Oman assessed by the Colored Progressive Matrices. *Mankind Quarterly* 53: 238-244.

Kura, K. (2013). Japanese north-south gradient in IQ predicts differences in stature, skin color, income and homicide rate. *Intelligence* 41: 512-516.

Lynn, R. (1979). The social ecology of intelligence in the British Isles. *British Journal of Social and Clinical Psychology* 18: 1-12.

Lynn, R. (1980). The social ecology of intelligence in France. *British Journal of Social and Clinical Psychology* 19: 325-331.

Lynn, R. (2010). In Italy, north-south differences in IQ predict differences in income, education, infant mortality, stature, and literacy. *Intelligence* 38: 93-100.

Lynn, R. (2012). North-south differences in Spain in IQ, educational attainment, per capita income, literacy, life expectancy and employment. *Mankind Quarterly* 52: 265-291.

Lynn, R. (2017a). Sex differences in intelligence: The developmental theory. *Mankind Quarterly* 58: 9-42.

Lynn, R. (2017b). Sex differences in intelligence: Reply to comments. *Mankind Quarterly* 58: 145-156.

Lynn, R. & Cheng, H. (2013). Differences in intelligence across 31 regions of China and their economic and demographic correlates. *Intelligence* 41: 553-559.

Lynn, R. & Irwing, P. (2004). Sex differences on the Progressive Matrices: A meta-analysis. *Intelligence* 32: 481-498.

Lynn, R. & Vanhanen, T. (2002). *IQ and the Wealth of Nations.* Westport, CT: Praeger.

Lynn, R. & Vanhanen, T. (2012). *Intelligence: A Unifying Construct for the Social Sciences.* London: Ulster Institute for Social Research.

Lynn, R., Allik, J., Pullmann, H. & Laidra, K. (2004). Sex differences on the Progressive Matrices among adolescents: Some data from Estonia. *Personality and Individual Differences* 36: 1249-1255.

Lynn, R., Sakar, C. & Cheng, H. (2015). Regional differences in intelligence, income and other socioeconomic variables in Turkey. *Intelligence* 50: 144-149.

NHS (2011). Genetic clues to age of first period. Retrieved from https://www.nhs.uk/news/genetics-and-stem-cells/genetic-clue-to-age-of-first-period/

Owen, K. (1992). The suitability of Raven's Progressive Matrices for various groups in South Africa. *Personality and Individual Differences* 13: 149-159. doi:10.1016/0191-8869(92)90037-P

Pesta, B., McDaniel, M. & Betsch, S. (2010). Towards an index of well-being for the fifty US states. *Intelligence* 38: 160-168.

Pietschnig, J. & Voracek, M. (2015). One century of global IQ gains: A formal meta-analysis of the Flynn effect (1909-2013). *Perspectives on Psychological Science* 10: 282-306.

Raven, J.C. (1941). Standardization of Progressive Matrices, 1938. *British Journal of Medical Psychology* 19: 137-150.

Raven, J. (2000). The Raven’s Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology* 41: 1-48.

Raven, J. (2008). *Standard Progressive Matrices and Mill Hill Vocabulary Scale*. London: Pearson.

Raven, J., Raven, J.C. & Court, J.H. (1985). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. London: H.K. Lewis & Co. Ltd.

Raven, J., Raven, J.C. & Court, J.H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Oxford: Oxford Psychologist Press.

Roivainen, E. (2012). Economic, educational and IQ gains in eastern Germany, 1990-2006. *Intelligence* 40: 571-575.

van der Linden, D., Dunkel, C. & Madison, G. (2017). Sex differences in brain size and general intelligence (*g*). *Intelligence* 63: 78-88.