Available online at www.sciencedirect.com
SCIENCE DIRECT
Intelligence 34 (2006) 121–139
INTELLIGENCE

Temperature, skin color, per capita income, and IQ: An international perspective

Donald I. Templer [a,*], Hiroko Arikawa [b]

[a] California School of Professional Psychology-Fresno, Alliant International University, United States
[b] Springfield, MO, United States

Received 5 February 2004; received in revised form 3 April 2005; accepted 7 April 2005
Available online 28 November 2005

Abstract

The impetus for our study was the contention of both Lynn [Lynn, R. (1991) Race differences in intelligence: A global perspective. *Mankind Quarterly*, 31, 255–296] and Rushton (Rushton [Rushton, J. P. (1995). *Race, evolution and behavior: A life history perspective*. New Brunswick, NJ: Transaction; Rushton, J. P. (1997). Race, intelligence, and the brain: The errors and omissions of the revised edition of S.J. Gould's the mismeasurement of man. *Personality and Individual Differences*, 23, 169–180; Rushton, J. P. (2000). Race, evolution, and behavior. *A life history perspective* (3rd edition). Port Huron: Charles Darwin Research Institute] that persons in colder climates tend to have higher IQs than persons in warmer climates. We correlated mean IQ of 129 countries with per capita income, skin color, and winter and summer temperatures, conceptualizing skin color as a multigenerational reflection of climate. The highest correlations were −0.92 (rho = −0.91) for skin color, −0.76 (rho = −0.76) for mean high winter temperature, −0.66 (rho = −0.68) for mean low winter temperature, and 0.63 (rho = 0.74) for real gross domestic product per capita. The correlations with population of country controlled for are almost identical. Our findings provide strong support for the observation of Lynn and of Rushton that persons in colder climates tend to have higher IQs. These findings could also be viewed as congruent with, although not providing unequivocal evidence for, the contention that higher intelligence evolves in colder climates. The finding of higher IQ in Eurasians than Africans could also be viewed as congruent with the position of Diamond (1997) that knowledge and resources are transmitted more readily on the Eurasian west–east axis.

Both Rushton (1995, 1997, 2000) and Lynn (1991) have pointed out that ethnic groups in colder climates score higher on intelligence tests than ethnic groups in warmer climates. They contend that greater intelligence is needed to adapt to a colder climate so that, over many generations, the more intelligent members of a population are more likely to survive and reproduce. Their temperature and IQ analyses have been descriptive rather than quantitative, however. In the present quantitative study, we predicted a negative correlation between IQ and temperature. We hypothesized that correlations would be higher for mean winter temperatures (January in the Northern Hemisphere and July in the Southern Hemisphere) than for mean summer temperatures. Skin color was conceptualized as a variable closely related to temperature. It is viewed by the present authors as a multigenerational reflection of the climates one's ancestors have lived in for thousands of years. Another reason to predict correlations of IQ with temperature and skin color is the product–moment correlation reported by Beals, Smith, and Dodd (1984) of 0.62 between cranial capacity and distance from the equator. Beals et al. based their finding on 20,000 individual crania

* Corresponding author.
E-mail address: donaldtempler@sbcglobal.net (D.I. Templer).

0160-2896/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2005.04.002

from every continent and representing 122 ethnically distinguishable populations. Jensen (1998) reasoned that natural selection would favor a smaller head with a less spherical shape because of better heat dissipation in hot climates. Natural selection in colder climates would favor a more spherical head to accommodate a larger brain and to have better heat conservation.

We used an index of per capita income-real gross domestic product (GDP) per capita to compare the correlations of income with IQ to those of temperature and skin color with IQ. There is a strong rationale for predicting a positive relationship between IQ and real GDP per capita. Common sense dictates that more intelligent populations can achieve greater scientific, technological, and organizational advancement. Furthermore, it is well established that conditions associated with poverty, such as malnutrition and inadequate prenatal/perinatal and other health care, can prevent the attainment of genetic potential. Lynn and Vanhanen (2002) did indeed find positive correlations between adjusted IQ and real GDP per capita of nations throughout the world. Their scatter plots vividly show that countries south of the Sahara Desert have both the lowest real GDPs per capita in the world and the lowest mean IQs in the world (in the 60s and 70s). The real GDP per capita in high-IQ countries is much more variable. For example, China and Korea have very high mean IQs but rather low real GDPs per capita. In this study, we considered only countries ($N = 129$) with primarily indigenous people—those with populations that have persisted since before the voyages of Christopher Columbus. It is acknowledged that there have been many migrations both before and after Columbus. However, the year 1492 has previously been used to define indigenous populations (Cavalli-Sforza, Menzoni, & Piazza, 1994).

1. Method

Mean IQs and real GDPs per capita of 129 countries were obtained from Lynn and Vanhanen (2002). Although these authors provided data for 185 countries, we considered only 129 countries because we wanted to restrict our research to indigenous peoples. Thus, we excluded Australia, New Zealand, countries in North and South America and the Caribbean, and very small islands with no data on the skin color. We also excluded Israel because people from many countries have settled there since the nation's inception in 1948. Lynn and Vanhanen's IQ means came from different tests administered in different eras in countries that differed in educational attainment. However, these authors adjusted for the Flynn effect—the improved performance on intelligence tests noted in recent decades. Moreover, 50 of the 55 calculated IQs used in our analyses were obtained with the Raven Progressive Matrices and/or the Cattell Culture Fair Test. Both instruments are devoid of educational and specific culture content. A further caveat is that Lynn and Vanhanen (2002) calculated IQs for 81 of 185 countries but estimated IQs for the other 104 countries, using mean IQs of neighboring countries. For the 129 countries used in the present study, 55 mean IQs were calculated and 74 were estimated by Lynn and Vanhanen. The similarity of the correlations for calculated and estimated IQs (Table 4) supports this aspect of Lynn and Vanhanen's methodology.

A physical anthropology source was used to obtain data on skin color (Biasutti, 1967). It should be noted, however, that physical anthropologists have traditionally assessed skin color inside the upper arm, which is affected only minimally by sun exposure. The source contains a map of the world with eight categories of skin color ranging from 1 (very light) to 8 (very dark). Because the map does not delineate the various countries of the world, three graduate students who were unaware of the purpose of our study independently determined the predominant skin color for each of the 129 countries. The word predominant was used because some countries had more than one skin color. The product–moment correlation coefficients between raters were 0.95, 0.95, and 0.93, suggesting very little subjectivity. For each country, the mean of the three skin color ratings was used.

Mean January high temperature, mean January low temperature, mean July high temperature, and mean July low temperature for all 129 countries were obtained from *Fodor's World Weather Guide* (Pearce & Smith, 1998). The guide covers many cities in larger countries but only one or two in small countries. It provided information for 10 Indian cities, 10 Chinese cities, and 11 Russian cities. For each country, we averaged the temperature information for all the cities covered by the guide. We defined winter temperatures as January temperatures in the Northern Hemisphere and July temperatures in the Southern Hemisphere. Summer temperatures were July temperatures in the Northern Hemisphere and January temperatures in the Southern Hemisphere.

2. Results

Table 1 contains the means and standard deviations of IQ, skin color, winter (January in Northern Hemisphere and July in Southern Hemisphere) mean high

temperature, summer mean low temperature, and 1998 mean real GDP per capita in US dollars. Table 2 contains the mean IQ, the skin color, real GDP per capita, and the four mean temperatures for all of the 129 countries included in the study. Table 3 contains the product–moment intercorrelations for the independent variables. Table 4 shows the product–moment correlation coefficients of IQ with all of the independent variables. It presents correlations for the 55 countries whose mean IQs were calculated by Lynn and Vanhanen, the 74 countries whose means were estimated by those authors, and all 129 countries. It is apparent that all of the correlations with skin color, winter temperatures, and real GDP per capita are significant. Mean winter temperatures correlated more strongly with IQ than mean summer temperatures. Moreover, the correlation of IQ with skin color, -0.92, was significantly ($p<0.001$) higher than the correlation of IQ with mean winter high temperature, significantly ($p<0.001$) higher than the correlation of IQ with mean winter low temperature, and significantly ($p<0.001$) higher than the correlation of IQ with real GDP per capita.

We performed multiple regression analysis with IQ as the dependent variable and the independent variables of skin color, mean winter high temperature, mean winter low temperature, mean summer high temperature, mean summer low temperature, and real GDP per capita. Real GDP per capita was the only independent variable that significantly ($p<0.001$) improved prediction beyond skin color. With this second variable, R rose from 0.917 to 0.922, and R^2 rose from 0.841 to 0.850.

In the statistical analyses with all 129 countries population of country was not controlled for. China with 1.2 billion persons was weighted the same as Estonia with 1.4 million persons. The correlations with IQ were again performed but controlling for 2001 population of country. The product–moment correlations controlling for population were almost identical to those not controlling for population. They were -0.92 for skin color, 0.64 for real GDP per capita, -0.76 for winter high, -0.66 for winter low, -0.31 for summer high, and -0.41 for summer low. The near identical correlations can be understood in terms of the zero order correlation ($r=-0.015$) between IQ and population.

An inspection of Table 2 suggests that the high correlations are in part a function of the large discrepancy between the Black African countries combined and the other countries combined in IQ, skin color, per-capita income, and temperature. Therefore, the correlations were performed separately for the 41 Black African nations and the other 88 nations. The 41 Black African countries are Angola, Benin, Botswana, Burkina Faso, Burundi, Cameron, Central African Republic, Chad, Congo (Brazzaville), Congo (Zaire), Cote d'Ivoir, Djibouti, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Malawi, Mali, Mozambique, Namibia, Niger, Nigeria, Rwanda, Senegal, Sierre Leone, Somalia, South Africa, Sudan, Swaziland, Tanzania, Togo, Uganda, Zambia, and Zimbabwe. Table 5 presents the product–moment and rank order correlations of IQ with the independent variables for 41 Black African countries and for the other countries. There appear to be two salient generalizations permitted by Table 4. One is that the correlations are much lower for the Black African countries. Such lower correlations are very understandable in view of the restricted variance in these countries. The standard deviations for the Black African and other countries respectively are 3.46 and 7.56 in IQ, 0.45 and 1.42 in skin color, $1769.37 and $8444.85 in real GDP, 5.94 and 10.84 in winter low temperature, 4.23 and 11.68 in winter high temperature, 3.8 and 5.39 in summer low temperature, and 3.87 and 5.71 in summer high temperature. The second generalization is that the pattern of correlations for the 88 other countries combined is rather similar to the pattern with all 129 countries.

Table 1
Means and standard deviations for all variables

	M	SD
IQ	84.47	12.5
Skin color	3.98	2.37
Real gross domestic product per capita	7024.51	7924.11
Celsius temperature		
Winter high	15.78	12.78
Winter low	6.68	11.28
Summer high	29.22	5.18
Summer low	19.46	4.98

3. Discussion

The negative correlations between IQ and temperature are congruent with the observations of Rushton and of Lynn that higher IQs are found in colder climates. The fact that the correlations between winter temperatures and IQ were higher than those between summer temperatures and IQ provide further support for their contentions. (The reasons for the correlations with winter high temperature being some-

Table 2
Calculated (cal) and estimated (est) IQ, skin color, real GDP per capita, and mean Celsius temperatures for all countries

Country	IQ		Skin Color	GDP	Population	Temperature			
	cal	est				Winter		Summer	
						Mean high	Mean low	Mean high	Mean low
Afghanistan		83	3.00	1200	26,813	8	−5	36	18
Albania		90	1.67	2804	3510	13	4	30	18
Algeria		84	4.33	4792	31,736	17	7	38	25
Angola		69	7.00	1821	10,366	23	13	27	19
Armenia		93	1.67	2072	3336	−2	−9	34	17
Austria	102		1.00	23,166	8151	0	−6	25	14
Azerbaijan		87	2.00	2175	7771	3	−3	32	17
Bahrain		83	4.00	13,111	645	20	14	37	29
Bangladesh		81	4.33	1361	131,270	26	13	31	26
Belarus		96	1.33	6319	10,350	−5	−11	24	14
Belgium	100		1.00	23,223	10,259	4	−1	21	12
Benin		69	7.00	867	6591	27	23	26	23
Bhutan		78	3.00	1536	2049	10	5	22	20
Botswana		72	7.00	6103	1586	24	5	31	18
Brunei		92	4.00	16,765	344	30	24	31	25
Bulgaria	93		1.67	4809	7707	4	−3	29	17
Burkina Faso		67	7.67	870	12,272	33	16	33	23
Burma		86	3.00	1199	41,995	28	14	30	24
Burundi		70	7.00	570	6224	29	17	28	19
Cambodia		89	5.00	1257	12,492	31	21	32	24
Cameroon		70	7.00	1474	15,803	30	21	27	20
Central African Republic		68	7.33	1118	3577	32	20	29	21
Chad		72	7.00	856	8707	32	13	38	23
China	100		2.00	3105	1,273,111	7	−3	29	21
Congo (Brazzaville)	73		6.67	995	2894	28	17	31	21
Congo (Zaire)	65		7.00	822		27	14	30	19
Cote d'Ivoire		71	6.33	1598	16,393	31	23	28	23
Croatia	90		2.00	6749	4334	12	6	29	21
Cyprus		92	2.00	17,482	763	16	7	35	22
Czech Republic	97		1.33	12,362	10,264	1	−5	24	14
Denmark	98		1.00	24,218	5353	3	−2	22	14
Djibouti		68	6.00	1266	461	29	23	41	31
Egypt	83		4.00	3041	69,537	20	10	35	23
Equatorial Guinea	59		6.00	1817	486	31	19	29	21
Eritrea		68	6.33	833	4298	26	14	29	18
Estonia		97	1.00	7682	1423	−4	−10	20	12
Ethiopia	63		6.67	574	65,892	24	9	22	12
Finland	97		1.00	20,847	5176	−5	−13	20	11
France	98		1.00	21,175	59,551	8	2	25	15
Gabon		66	7.00	6353	1221	28	20	31	23
Gambia		65	8.00	1453	1411	31	15	30	23
Georgia		93	2.00	3353	4989	7	−1	31	19
Germany		102	1.00	22,169	83,030	2	−3	23	13
Ghana		71	7.00	1735	19,894	33	22	28	23
Greece		92	2.00	13,943	10,624	11	5	32	21
Guinea	66		7.67	1782	7614	31	22	28	22
Guinea–Bissau		66	7.33	616	1316	29	17	31	24
Hong Kong	107		2.00	20,763	7211	18	13	31	26
Hungary	99		1.00	10,232	10,106	1	−5	28	16
Iceland		98	1.00	25,111	278	2	−2	14	9
India	81		6.33	2077	1,029,991	19	10	29	22
Indonesia	89		4.67	2651	228,438	30	23	30	23
Iran	84		3.00	5121	66,129	13	3	38	25
Iraq	87		3.33	3197	23,332	17	6	42	26
Ireland	93		1.00	21,482	3841	8	2	19	12

Table 2 (*continued*)

Country	IQ		Skin Color	GDP	Population	Temperature			
	cal	est				Winter		Summer	
						Mean high	Mean low	Mean high	Mean low
Italy	102		1.67	20,585	57,680	11	4	29	20
Japan	105		2.00	23,257	126,772	5	−3	27	18
Jordan		87	3.00	3347	5153	12	4	32	18
Kazakhstan		93	2.00	4378	16,731	−7	−15	27	16
Kenya	72		6.67	980	30,766	25	16	28	18
Korea, North		104	2.00	3000	21,968	−1	−11	28	20
Korea, South	106		2.00	13,478	47,904	3	−6	28	22
Kuwait		83	4.00	25,314	2042	16	9	39	30
Kyrgyzstan		87	2.00	2317	4753	−1	−10	30	17
Laos		89	4.33	1734	5636	28	14	31	24
Latvia		97	1.00	5728	2385	−4	−10	22	11
Lebanon	86		3.67	4326	3628	14	6	31	19
Lesotho		72	7.00	1626	2177	16	1	30	16
Liberia		65	7.00	1200	3226	30	23	27	22
Libya		84	4.33	6697	5241	17	9	29	22
Lithuania		97	1.33	6436	3611	−5	−11	23	12
Luxembourg		101	1.00	33,500	443	3	−1	23	13
Macedonia		93	1.67	4254	2046	5	−3	31	15
Malawi		71	7.00	523	10,548	23	7	27	17
Malaysia	92		4.67	8137	22,229	29	21	30	21
Mali		69	6.00	681	11,009	32	15	36	24
Mauritania		74	5.00	1563	2747	29	14	32	23
Moldova		95	2.00	1947	4432	−1	−8	27	16
Mongolia		98	2.00	1541	2655	−19	−32	22	11
Morocco	85		2.67	3305	30,645	18	6	33	18
Mozambique		72	7.00	782	19,371	26	14	32	22
Namibia		72	6.67	5176	1798	21	7	26	16
Nepal	78		4.33	1157	25,284	18	2	29	20
Netherlands	102		1.00	22,176	15981	5	0	22	13
Niger		67	7.00	739	10,355	34	14	34	23
Nigeria	67		7.00	795	126,636	31	19	29	22
Norway	98		1.00	26,342	4503	−2	−7	17	10
Oman		83	5.00	9960	2622	25	19	36	31
Pakistan		81	3.67	1715	144,617	20	7	38	27
Philippines	86		4.00	3555	82,842	30	23	31	24
Poland	99		1.00	7619	38,634	0	−5	24	14
Portugal	95		2.00	14,701	10,066	13	6	27	16
Qatar	78		4.00	20,987	769	22	13	38	29
Romania	94		2.00	5648	22,364	2	−6	28	16
Russia	96		2.00	6460	145,470	−13	−20	23	13
Rwanda		70	7.00	660	7313	26	12	25	14
Saudi Arabia		83	4.00	10,158	22,757	25	13	39	26
Senegal		65	7.67	1307	10,285	26	18	31	24
Sierre Leone	64		7.00	458	5427	29	24	28	23
Slovakia	96		1.33	9699	5415	1	−5	26	15
Slovenia	95		1.00	14,293	1930	2	−4	27	14
Somalia		68	7.00	1000	7489	30	22	35	27
South Africa		72	6.67	8488	43,586	19	5	27	16
Spain	97		2.00	16,212	40,038	12	6	28	17
Sri Lanka		81	6.00	2979	19,409	25	18	27	21
Sudan	72		6.67	1394	36,080	32	18	37	24
Swaziland		72	7.00	3816	1104	19	6	25	15
Sweden	101		1.00	20,659	8875	−2	−7	22	13
Switzerland	101		1.00	25,512	7283	1	−5	21	14
Syria		87	3.33	2892	16,729	11	2	39	22

(continued on next page)

Table 2 (*continued*)

Country	IQ		Skin Color	GDP	Population	Temperature			
	cal	est				Winter		Summer	
						Mean high	Mean low	Mean high	Mean low
Taiwan	104		3.00	13,000		21	15	32	25
Tajikistan		87	2.67	1041	6579	−1	−10	30	17
Tanzania	72		7.00	480	36,232	28	16	29	21
Thailand	91		3.67	5456	61,798	30	17	32	24
Togo		69	7.00	1372	5153	31	22	27	23
Tonga	87		5.00	3000	104	30	24	29	23
Tunisia		84	3.00	5404	9705	15	5	34	21
Turkey	90		2.00	6422	66,494	6	−2	28	16
Turkmenistan		87	2.33	2550	4603	3	−2	34	23
Uganda	73		7.67	1074	23,986	26	15	24	14
Ukraine		96	1.67	3194	48,760	−1	−8	27	16
United Arab Emirates		83	4.00	17,719	2407	23	12	38	28
United Kingdom	100		1.00	20,336	59,648	6	2	19	12
Uzbekistan		87	2.00	2053	25,155	3	−6	33	18
Vietnam		96	4.00	1689	79,939	25	18	33	25
Yemen		83	6.00	719	18,078	28	23	36	29
Yugoslavia		93	2.00	4000		6	0	31	16
Zambia	77		7.67	719	9770	24	8	26	17
Zimbabwe	66		7.00	2669	11,365	21	7	27	16

what higher than with winter low temperature are not clear. It is possible that in the process of evolution the active coping with the environment took place more during the warmer daylight hours.) It should be pointed out, however, that the correlation of IQ with GDP per capita is of about the same magnitude as the correlation of IQ with winter temperature. It should further be noted that the correlations of IQ with skin color and temperature are congruent with the correlation of 0.62 between cranial capacity and distance from the equator reported by Beals et al. (1984). Skin color is both more theoretically related to climate than to income and more rooted in biology than either temperature or income. The most salient finding is the very high (for the social and behavioral sciences) correlation of −0.92 between skin color and IQ score. It should be noted, however, that we conceptualized skin color as a climatic variable. It would be absurd to suggest that a lighter complexion makes people more intelligent or that higher intelligence lightens the skin. A possible reason why skin color correlated more highly than temperature with IQ is that it is a multigenerational reflection of climatic history. It takes thousands of years for skin color to change through evolution. For example, desert Amerindians are lighter than Negroid persons south of the Sahara Desert because they have not lived in a hot climate for as long. Mongoloid persons moved from Siberia to the Americas relatively recently. Nevertheless, it should be acknowledged that the present findings provide more support for the contention of Rushton and Lynn that higher IQs are found in colder climates than their postulated evolutionary processes. The formulation of any postulated evolutional process involves hundreds if not thousands of facts and relationships. The correlations of the present study are consistent with the evoluational postulates of Rushton and Lynn but do not provide definitive evidence.

Table 3
Intercorrelations of independent variables—temperature, skin color, and real GDP per capita

	Winter high	Winter low	Summer high	Summer low	Skin color
Winter high					
Winter low	0.96*				
Summer high	0.45*	0.41*			
Summer low	0.68*	0.70*	0.79*		
Skin color	0.85*	0.75*	0.30*	0.47*	
Real GDP per capita	−0.43*	−0.33*	−0.36*	−0.30*	−0.60*

* $p<0.001$.

It could be argued that the Lynn/Rushton evolutionary/genetic position provides only one of several plausible explanations for the present findings. One of these is the formulation of Diamond (1997) who categorically rejects genetic determination. Diamond maintained that

Table 4
Product–moment and rank order correlations of independent variables with mean IQ of country

Independent variables	IQ					
	Calculated countries ($N=55$)		Estimated countries ($N=74$)		All countries ($N=129$)	
	r	rho	r	rho	r	rho
Skin color	−0.90**	−0.84**	−0.93**	−0.92**	−0.92**	−0.91**
Real GDP per capita	0.74**	0.84**	0.44**	0.59**	0.63**	0.74**
Celsius temperature						
Winter high	−0.74*	−0.73**	−0.78**	−0.80**	−0.76**	−0.76**
Winter low	−0.63**	−0.66**	−0.69**	−0.71**	−0.66**	−0.68**
Summer high	−0.30*	−0.37*	−0.22	−0.18	−0.31**	−0.33*
Summer low	−0.31*	−0.34**	−0.41**	−0.46**	−0.40**	−0.45**

* $p<0.05$.
** $p<0.001$.

it is geography and not biology that accounts for the earlier and higher levels of technological development in some parts of the world than in other parts. His assumption that geographically isolated regions cannot benefit from innovations in other regions is self-evident. Diamond methodically explained why only a small minority of plants and animals can be domesticated, and that inventions are dependent upon prior inventions. The geography of Eurasia facilitates the diffusion of ideas, knowledge, animals, plants and technology on its west–east axis. The greater similarity of climate along a west–east axis is also an advantage in contrast to the north–south axis of the Americas and Africa. The Sahara Desert is a major geographical impediment to the diffusion of knowledge and resources in Africa. It is likely that the Diamond's conceptualization contributes to some but not all of the big picture provided by the present findings and assorted research. It does not mesh well with the correlation of 0.62 between cranial capacity and distance from the equator reported by Beals et al. (1984). There is probably no one conceptual contribution that explains 100% of the present findings.

Table 5
Product–moment and rank order correlations will mean IQ for the 41 Black African countries and the 88 other countries

Independent variables	IQ			
	41 Black conntries		88 other countries	
	r	rho	r	rho
Skin color	0.06	−0.12	−0.74**	−0.77**
Real GDP per capita	0.14	0.08	0.53**	0.56**
Celsius temperature				
Winter high	−0.32*	−0.32*	−0.56**	−0.59**
Winter low	−0.37*	−0.38*	−0.47**	−0.53**
Summer high	−0.01	−0.10	−0.61**	−0.66**
Summer low	−0.25	−0.34*	−0.57**	−0.62**

* $p<0.05$.
** $p<0.001$.

The rationale for conceptualizing skin color as a climate-related variable is supported by the fact that, in our multiple regression analysis, temperature did not improve prediction beyond that provided by skin color alone. Nevertheless, it cannot be denied that skin color is also a biological variable. In our opinion, at least on an international level, it now deserves to be categorized with brain size and musculoskeletal traits (Rushton & Rushton, 2003) as a biological correlate of intelligence. The correlation between IQ and skin color in individual countries is probably much weaker because skin color does not vary as much in a single country as globally. Also, it should be borne in mind that the correlations are based upon group means and not individuals and so result in higher orders of magnitude than would otherwise be the case. It is recognized that malnutrition and tropical diseases can affect the brain, and inadequate medical care can lower intelligence in poor countries. However, the finding that IQ correlates more highly with skin color than with income could be viewed as congruent with the Lynn/Rushton conceptualization.

The Black African vs. rest of the world discrepancies could be viewed as consistent with the review of Jensen (1998) on genetic distance between ethnic groups. Jensen pointed out that the genetic distance on a number of different biological variables between Caucasoids and Mongoloids is less than that between Negroids and the other two groups. These biological variables include brain size, blood groups, enzymes, and proteins involved in the immune system. Jensen noted that such a Negroid vs. non-Negroid distinction is also obtained on the physical characteristics that physical anthropologists traditionally use in categorizing the "races." Nevertheless, the composite of the present evidence and

genetic distance findings does not permit unequivocal statements about the genetic determination of the present findings. The great poverty of Africa produces conditions that can unquestionably lower IQ such as malnutrition, disease, inferior prenatal care, and inferior perinatal care. The incidence of epilepsy is very high in Africa, and this has been attributed to malnutrition, parasitic and other brain infections, febrile convulsions, and birth injuries (Osantokun et al., 1987).

It could be argued that there are methodological limitations with respect to the mean IQs provided by Lynn and Vanhanen (2002). One possible limitation is that for 74 of the countries used in the present study the IQs were estimated on the basis of IQs in neighboring countries. The present authors maintain that the assumption of neighboring countries tending to have similar IQs is not entirely arbitrary and unjustifiable. We use the analogy of height because height and IQ have similar determinations such as genetics, nutrition, and health care. It would seem more reasonable to predict height of Norwegians from the height of Swedes than from the height of Italians. Furthermore, the fact that the correlations for the calculated IQ countries and estimated IQ countries are similar lends credulence to the legitimacy of the Lynn and Vanhanen procedure for estimating mean IQs. Another possible limitation is that the Lynn and Vanhanen means are based on different tests administered in different eras and in countries that differ in average educational attainment. However, Lynn and Vanhanen made adjustments for the "Flynn effect," an increase in intelligence test performance in recent decades (Flynn, 1987). And, for 50 of the 55 countries in which IQ was calculated and used in the present study, the Raven Progressive Matrices and/or the Cattell Culture Fair Test were used. Although both of these instruments appear to be void of educational and specific culture content, it cannot be assumed that they are equally effective in measuring intelligence around the world. The fact that skin color is not uniform within countries (as displayed in Biasutti, 1967), could also be seen as a methodological limitation. However, the very high inter-rater reliability, combined with the high correlations with skin color, indicate the effects of this limitation are rather small. Furthermore, the positive correlation between skin color and temperature provide evidence for the validity of the skin color map employed. The fact that the *N* for some of the mean IQ's of the countries provided by Lynn and Vanhanen are below optimal constitutes an additional limitation. Measurement instrument limitations ordinarily attenuate rather than inflate correlations. Our correlations are high.

Acknowledgments

We thank Joan Ensher for library research, anthropologist, Roger LaJeunesse for helpful suggestions.

References

Beals, K. L., Smith, C. L., & Dodd, S. M. (1984). Brain size, brain morphology, climate, and time machines. *Current Anthropology*, *25*, 301–328.

Biasutti, R. (1967). *Le Razze e i Popoli Della Terra*. Torino, Italy: Unione pipografiza-Editrice Torinese.

Cavalli-Sforza, L. L., Menzoni, P., & Piazza, A. (1994). *The history and geography of human genes*. Princeton: Princeton University Press.

Diamond, J. (1997). Guns, germs and steel. New York: W.W. Norton & Company.

Flynn, J. R. (1987). Massive IQ gains in 14 nations. *Psychological Bulletin*, *101*, 1171–1191.

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

Lynn, R. (1991). Race differences in intelligence: A global perspective. *Mankind Quarterly*, *31*, 255–296.

Lynn, R., & Vanhanen, T. (2002). *IQ and the wealth of nations*. Westport: Praeger Publishers.

Osantokun, B. O., Adeuja, A. O., Nottidge, V. A., Bademosi, O., Olumide, A., Inge, O., et al. (1987). Prevalence of the epilopsies in Nigerian Africans: A community-based study. *Epilepsia*, *28*, 272–279.

Pearce, E. A., & Smith, C. G. (1998). *Fodor's world weather guide*. New York: Random House.

Rushton, J. P. (1995). *Race, evolution and behavior: A life history perspective*. New Brunswick, NJ: Transaction.

Rushton, J. P. (1997). Race, intelligence, and the brain: The errors and omissions of the revised edition of S.J. Gould's the mismeasure of man. *Personality and Individual Differences*, *23*, 169–180.

Rushton, J. P. (2000). *Race, evolution, and behavior. A life history perspective* (3rd edition). Port Huron: Charles Darwin Research Institute.

Rushton, J. P., & Rushton, W. W. (2003). Brain size, IQ, and racial group differences: Evidence from musculoskeletal traits. *Intelligence*, *31*, 139–155.

Discussions

Comments on correlations of IQ with skin color and geographic–demographic variables

Abstract

A large number of national and geographic population samples were used to test the hypothesis that the variation in mean values of skin color in the diverse populations are consistently correlated with the mean measured or estimated IQs of the various groups, as are some other physical variables,

known as an ecological correlation. Straightforward statistical analyses clearly bear out the hypothesis, showing a significant positive ecological correlation between lightness of mean skin color and mean IQ across different populations. The main limitation of such a study design is that correlations obtained from this type of analysis are completely non-informative regarding any causal or functional connection between individual differences in skin pigmentation and and individual differences in IQ, nor are they informative regarding the causal basis of the correlation, e.g., simple genetic association due to cross-assortative mating for skin color and IQ versus a pleiotropic correlation in which both of the phenotypically distinct but correlated traits are manifested by one and the same gene.

The simple, clear-cut design and statistical analysis in Templer's and Arikawa's (in press) study is entirely conventional, not unlike innumerable other studies published in this and many other refereed psychological and sociological journals. The authors straightforwardly test a hypothesis put forth by other researchers in attempting to explain the observed cognitive racial differences in terms of evolutionary psychology. The main research question is whether a large sample of estimated mean IQs in a great many countries around the world are correlated with the mean temperatures in those regions. And because differences in the darkness of skin color also vary with the mean temperature different climatic regions, the next question is whether skin color, too, is related to IQ. The data are presented, the zero-order correlations calculated, and the hypothesized IQ × skin color correlation is statistically borne out.

The collection of mean IQs and estimated IQs in the various countries (whether *N*-weighted or not) almost certainly fall short of the degree of reliability and validity attainable with psychometric tests administered under laboratory conditions to carefully selected representative national population samples tested under virtual laboratory conditions. Nevertheless the study's rough and ready pick-up mental test data, probably because it was aggregated over some number of varied samples in each country, evidently has sufficient overall precision to show significant and meaningful *ecological correlations* with the other variables in the study. (So-called *ecological correlations* are based on the means of aggregated data, such as the data used in the present study. They are widely used in epidemiology, a branch of medical research, and they also have valuable applications in the behavioral sciences Lubinski & Humphreys, 1996). Ecological correlations are not necessarily less valid than correlations based on individuals' test scores as the unit of measurement, unless it can be successfully argued that, in a given set of data, the nominal ecological variable (e.g., various nations' mean IQs) represents an essentially different latent trait from that represented by test scores based on individuals. It is actually around this important consideration that the crux of the causal argument involving group differences exists between strict environmentalists and most hereditarians, who claim that some proportion of the total between-groups variance is genetic (e.g., Rushton & Jensen, 2005). The aggregation of individual test scores to form group means serves to "average-out" the unique and irrelevant sources of variance among individuals, thereby abstracting and highlighting what is presumably one and the same common factor reflected by the test scores of individuals. This culling out the irrelevant variance unrelated to the latent trait of interest is typically a basic, though not explicit, assumption of most uses of ecological correlations. The use of ecological correlations often calls for more detailed justification than is typically provided in the study of national or other group differences.

My preferred methodology for researching the meaning of correlations between phenotypically distinct variables, such as IQ and skin color, for example, begins with data that is as close as possible to the basic units of its phenotypic measurement in individuals. But to insist on that condition would be to insist that Templer and Arikawa should carry out an altogether different study from the one they have presented here, which is entirely above board in the details of its method, so typical readers of *Intelligence* can make their own subjective evaluations of the study's scientific importance in light of the data's limitations as explicitly announced by the authors. Their analysis yields an interesting set of findings, in the form of several ecological correlations. These could suggest other hypotheses concerning the actual cause of the correlation between IQ and skin color. The cause of the obvious correlation between skin pigmentation and climate (or temperature) is now well understood and is scarcely controversial. It is the relationship between skin color and IQ within as well as between the major racial groups that is of most interest.

Some years ago I reviewed the then total empirical literature on the IQ × skin color correlation (Jensen, 1973, pp. 222–224). The evidence from eighteen published studies indeed leaves very little doubt of a significant correlation *within* the racially hybrid population of African-Americans, in which today about 25% of its genes derive from Caucasian-European ancestors. The overall average of the reported IQ × -

skin color correlations is about +0.20, ranging from +0.12 to +0.30. (The correlations are positive because lighter degrees of pigmentation were assigned higher ratings). These correlations most probably reflect primarily the considerable genetic heterogeneity of a relatively recent hybrid population that has not yet interbred for enough generations to allow all genetic linkages between the genes affecting skin color and the genes affecting cognitive abilities to break up sufficiently to attain genetic equilibrium (i.e., when the Hardy-Weinberg law applies to all of the gene loci involved in the correlated traits). Templer and Arikawa emphasize that they regard skin color only as a climatic variable, a multigenerational reflection of climatic history. And this may well be theoretically adequate for their present purpose. But we should not let it mislead us to dismiss completely other possible, and presently causal, connections between skin color and IQ—an idea the authors, perhaps too cautiously, called "absurd." This stance overlooks the probability of the genetic phenomenon of *pleiotropy* acting as at least a partial cause of the IQ × skin color correlation in present day populations. (Pleiotropy is the condition of a single gene having two or more phenotypically quite different effects. For example a single gene could affect both IQ and skin color.) The main theory invoked to explain the IQ × skin color correlation is that the many millennia of successive migrations out of Africa into Europe and Asia, with their colder and often more severe climates, selected simultaneously both for skin color and for certain cognitive abilities, especially *g*. While both factors were also controlled by separate genes acting independently and would gradually approach Hardy-Weinberg equilibrium, it is also highly likely that in the very long course of this massive selection process even a rare genetic mutation resulting in *pleiotropic* genes that affect *both* pigmentation and *g* would have occurred and then multiplied in frequency because of their joint selective advantage in different climates.

The pleiotropy hypothesis makes sense in terms of evolutionary genetics. But can we empirically reject this pleiotropy hypothesis? After all, the possibility of outright empirical rejection of a hypothesis is the Popperian criterion of scientific argumentation. I do think it is possible to meet this criterion. I propose that it can be done by determining whether the IQ × skin color correlation is what I have elsewhere termed an *intrinsic* correlation, as contrasted with an *extrinsic* correlation (Jensen, 1980; Jensen & Sinha, 1993). The presence of an extrinsic correlation in the absence of an intrinsic correlation rules out pleiotropy. The methodology of making this distinction has been applied to the correlation of IQ with physical stature (extrinsic), of IQ with head size (intrinsic), and of IQ with myopia (intrinsic) (Jensen, 1980; Jensen & Johnson, 1994; Cohn, Cohn, & Jensen, 1988). An *extrinsic* correlation between variables X and Y is one in which the absolute value of $r_{XY}>0$ in the population and is not >0 within families (i.e., within full sibships). An intrinsic correlation is one for which the absolute value of $r_{XY}>0$. both in the *population* and *within* families. Although every pair of full siblings (including DZ twins) has exactly the same unique ancestral genealogy, the members of each pair differ in the particular selection of the parental genes they inherit at conception. An individual who inherits a pleiotropic gene manifests both of its phenotypic effects, such as lighter pigmentation and higher IQ, as would be hypothesized in the case of these two variables. All of the IQ × skin color correlations reported in the literature are entirely population correlations, hence they are not informative regarding pleiotropy. But with a reasonably large sample of full sibling pairs it would be possible to rule out pleiotropy. It would be ruled out if no statistically significant within-families correlation were found between siblings' IQs and the siblings' values on a linear index of skin pigmentation as objectively measured by one of the standard procedures used in physical anthropology. Pleiotropy implies that within each sibling pair the individual having the higher IQ would also more frequently have the lighter skin color. If this is not found to be the case, the pleiotropic hypothesis would have to be rejected. But if, on the other hand, the IQ × skin color correlation turns out to be pleiotropic, and if this result can be adequately replicated, it would constitute a key item of evidence for the co-evolution of IQ (or more specifically *g*) and skin color. Unless geneticists can find sufficient fault with this line of reasoning as to render the proposed study scientifically worthless or technically unfeasible, I would hope that such a study will soon be forthcoming.

References

Cohn, S. J., Cohn, C. M. G., & Jensen, A. R. (1988). Myopia and intelligence: A pleiotropic relationship? *Human Genetics*, *80*, 53–58.

Jensen, A. R. (1973). *Educability and group differences*. London: Methuen (New York: Harper & Row).

Jensen, A. R. (1980). Uses of sibling data in educational and psychological research. *American Educational Research Journal*, *17*, 153–170.

Jensen, A. R., & Johnson, F. W. (1994). Race and sex differences in head size and IQ. *Intelligence*, *18*, 309–333.

Jensen, A. R., & Sinha, S. N. (1993). Physical correlates of human intelligence. In P. A. Vernon (Ed.), *Biological approaches to the study of human intelligences* (pp. 139–242). Norwood, NJ: Ablex.

Lubinski, D., & Humphreys, L. G. (1996). Seeing the forest from the trees: When predicting the behavior or status of groups, correlate means. *Psychology, Public Policy, and Law*, *2*, 363–376.
Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law*, *1*, 235–294.
Templer, D. I., & Arikawa, H. (in press). Temperature, skin color, per capita income, and IQ: An international I perspective. *Intelligence*.

Arthur R. Jensen
University of California, Berkeley, United States
E-mail address: nesnejanda@aol.com.

doi:10.1016/j.intell.2005.04.003

Sorry, wrong numbers: An analysis of a study of a correlation between skin color and IQ

Abstract

We argue that the report by Templer and Arikawa contains misleading conclusions and is based upon faulty collection and analysis of data. The report fails to hold up for quality of data, statistical analysis, and the logic of science.

Templer and Arikawa (this issue) report a −0.9 correlation between mean national intelligence test scores and an index of skin color running from 1 (white) to 8 (extremely dark). They interpret this correlation as consistent with arguments by Lynn (1991), Rushton (1995), and Jensen (1998) that both skin color and intelligence are largely biologically determined variables, and that the two covary. Templer and Arikawa further stress an argument, originally due to Rushton, that the greater challenge of living in higher latitudes, plus the paucity of sunlight, favors the evolution of both superior mental capacity and lighter skin color. (One might equally argue that the greater challenge of living in equatorial regions, such as fending off myriad parasitic diseases, should render equatorial people more intelligent.) The authors do admit that there are non-biological explanations, such as Diamond's (1997) argument that the Earth's geography is such that the flow of ideas and technology (and hence prosperity) is easier along the Eurasian East–West axis than the North–South axis, and that societies where there is a constant interchange of ideas are likely to produce more analytic, enquiring minds.

Here we will maintain that the Templer and Arikawa data collection and analyses are seriously flawed. Even if their methods were technically adequate and if the claimed correlations existed, the correlations would be uninterpretable and hence of no scientific value.

Templer and Arikawa's variables are a national intelligence estimate, taken from Lynn and Vanhanen's (2002) analysis of the relation between estimated mean national IQ and gross domestic product per capita (GDPC), an estimate of "preponderant" skin color for the country in question, and mean winter and summer temperatures in the national capital. The data are provided in Templer and Arikawa's Table 1, which also provides the 2001 population of each country. The IQ and GDPC data were taken from Lynn and Vanhanen (2002).

We shall deal with three variables—IQ scores, the skin color index, and population. We shall also distinguish between scores on a putative test of intelligence (throughout, IQ) and intelligence as a concept, viewed here as individual differences in mental competence that can influence a person's success in life.

1. Objections based on the quality of the data

The population data are, of course, no more suspect than any routinely collected census data. The IQ and skin color indices are more suspect.

As Templer and Arikawa accepted the Lynn and Vanhanen data at face value, any weakness in the Lynn and Vanhanen data is inherited by the Templer and Arikawa study. The Lynn and Vanhanen data set is far from ideal.

Lynn and Vanhanen's IQ data were based on reports from a variety of studies in 81 countries. Virtually none of the original studies claimed to be based on national samples. For example, several of their data sets were what the original authors described as standardization samples for the Raven Matrices tests. These standardization samples were by no means population samples. In some cases they consisted of school children in a single town or city. In general, when Lynn and Vanhanen had two or more samples within the same country, they averaged them without weighting for sample size. Even if we accept the validity of the various IQ tests across cultures, a point to which we will return, estimates obtained in this way strike us as being, on statistical grounds alone, inadequate estimates of national IQ.

In order to develop a larger sample of countries, Lynn and Vanhanen estimated IQ scores for a further

104 countries, extrapolating from the 81 observed data points. Their extrapolation method was based on the assumption that countries that are geographically close to each other are likely to have populations with similar IQs. Such an assumption would suggest, for example, that average IQs in Mexico and the United States are likely to be substantially more similar than average IQs in, say, the United Kingdom and the US, despite the substantial shared language and cultures of the UK and the US.

Templer and Arikawa decided to exclude countries that had been subject to immigration pressures since 1500, which thus excluded all the Western Hemisphere, Australia and New Zealand. This led them to select 55 countries for which Lynn and Vanhanen actually had data and 74 for which the data were estimated in what can only be viewed as a suspect way.

We realize that it would be extremely difficult to collect national samples of IQs on a worldwide basis. However, we do not regard the fact that the right data are hard to obtain as an excuse for drawing conclusions based on flawed data, simply because the flawed data are easily available.

Whatever the validity of the geographic technique of extrapolation, the practice of imputing 104 data points from 81 observations is, in our view, indefensible. Both Templer and Arikawa and Lynn and Vanhanen point out that their correlations for the observed and imputed data are virtually identical. We strongly suspect that this is because the technique of imputing 104 points from 81 observables virtually guarantees that relationships existing within the 81 real data points carry over to the imaginary ones.

In order to estimate skin color Templer and Arikawa first consulted reports in a 1967 anthropology book, but found that it did not give estimates for skin color by nation. Therefore they used the following procedure, which we quote:

"...three graduate students who were unaware of the purpose of our study independently determined the predominant skin color for each of the 129 countries."

What Templer and Arikawa meant by "determined" is unclear, but in an open presentation at the International Society for Intelligence Research in 2004, Templer, in response to a question, indicated that it was the graduate students' opinions, and that insofar as he knew, the graduate students had no particular first-hand experience with any of these countries.

Templer and Arikawa then say: "The product–moment correlation coefficients between raters were 0.95, 0.95, and 0.93, suggesting very little subjectivity." Inter-rater agreement measures, such as the correlations reported by Templer and Arikawa, are indices of the reliability of a measurement, not of its validity. Therefore, the issue of subjectivity has not been addressed. All we know is that on this topic three graduate students think alike. They may share the same implicit theories, prejudices, erroneous preconceptions, or whatever. For example, the fact that three judges from the Salem witch trials might have shared the same views as to which of the accused were witches did not make the accused witches.

The "skin color" referred to in the paper is the skin color that graduate students who had not visited the relevant countries, and, for all we know, had no relevant knowledge whatsoever of those countries, *think* is predominant in each country! If Templer and Arikawa's paper had been entitled "IQ is correlated with graduate students' stereotypes of skin color across nations," we might not have objected so vehemently as we do here (although we would like to see a rather larger and more diverse sample of graduate students). The paper would then become a paper about stereotypes rather than biological variables. In fact, that is what we think it is.

Why is this important? It is well known that many countries with dark-skinned people are desperately poor. If one has the belief that poverty is associated with low levels of mental competence (without in any way implying causation), or even that being "non-Western" is associated with low scores on Western-developed measures of mental competence–laying aside for the moment whether or not those measures are meaningful–the correlation Templer and Arikawa report would be produced. But this correlation would be based upon the beliefs, or implicit theories, of the raters, regardless of what the facts on the ground are.

We add that it would not be necessary for judges to behave consciously in the manner just described in order to produce this effect. Highly publicized research on implicit associations (Nosek, Banaji, & Greenwald, 2002) has shown that people's behavior can be influenced by the experience of statistical associations between ideas (e.g., media reports of dark-skinned people being associated with poverty situations, or of their having low intelligence test scores) without the people's being aware that their actions are driven by the association.

Psychological research often relies on ratings by judges. A basic principle in the design of such studies is that every effort should be made to ensure that judges' preconceptions do not influence the conclusions to be drawn from analysis of the ratings. We

believe that the Templer and Arikawa method for obtaining skin color indices would be unacceptable even in an undergraduate paper. We are more than surprised to find it acceptable in a refereed journal.

We also question the notion of "predominant color." This notion may have some validity for a small homogeneous country, in a situation where the national boundary coincides with historic ethnicity. It may be possible to assign a predominant color to the 104,000 residents of Tonga; we don't know. But when we deal with large, multi-ethnic nations such as China or India, the concept appears to us to be ridiculous. There are few large, ethnically homogeneous countries left. One has only to visit countries that may have once been homogeneous, such as England or France or Germany, to see how great their current diversity is.

The same objection applies to the IQ data. When dealing with dramatically different levels of health, educational opportunity and economic development within the country, the concept of national IQ is meaningless without carefully designed probability samples of the population. What is the national IQ of the US, for example? There is tremendous range, and averages may vary widely across different parts of the country. Does the average reflect anything in general about the US or its citizens? We think not.

In sum, whatever the relationship is between "true intelligence" and "true skin color," the measures offered in this study were fatally flawed.

2. Objections based upon the statistical analysis

Templer and Arikawa treat each nation as a single data point, regardless of size. The correlation between the measures should have been established by allowing for the size of the nation. These effects could be dramatic, as the ratio of the largest population in their data set (China) to the smallest (Tonga) is over 12,000 to 1.

Templer and Arikawa claim that they controlled for size by "adjusting statistically." It is not clear what they mean by this, but from their text it appears to mean that they computed the partial correlation between IQ and skin index, after allowing for correlations associated with population size.

This procedure is obviously inappropriate. The issue is not whether size predicts either of the other indices, but what the nation size–skin color correlations are when each data point is weighted by the size of the country. In fact, when we recalculated the correlations for the 53 nations in Templer and Arikawa's data set for which IQ data exists, the correlation is −0.91, virtually identical with Templer and Arikawa's report. However, the correlation is extremely sensitive to the assignment of skin color indices to large countries. To illustrate, Templer and Arikawa assign India a skin color index of 6.33, a value shared only with the Ivory Coast. They assign China an index value of 2, along with (among others) Russia and Croatia, and, surprisingly, lighter than the 3 assigned to Taiwan. Suppose that we drop the Indian index to 4.0 (the average of Pakistan, on the West, and Bangladesh, on the East), and raise China to 3, the value assigned to Taiwan. The correlation now drops to −0.81, a drop of about 15% in variance accounted for, by minor changes in just two data points.

We do not argue that our estimates are correct or incorrect. The point of the exercise is solely to illustrate that Templer and Arikawa's analysis is quite sensitive to the assignment of predominant skin indices to large countries. These are the data points most suspect, both because of the arbitrariness of Templer and Arikawa's numbers (how many of our readers would care to distinguish between Taiwanese and mainland Chinese on the basis of skin color?) and because the large countries are the very countries where the concept of any one predominant color is most suspect.

When Templer and Arikawa computed correlations between estimates of IQ and skin color, they implicitly assumed that if two countries have the same mean IQ score they are, in some sense, inhabited by equivalently intelligent people. More technically, they assumed what is known as *full-score comparability*, that is, that a numerical IQ score in one culture means the same as the identical number in another culture (Van de Vijver and Leung, 1997). The issue of cross-cultural comparability of IQ scores has been the subject of a very large research effort, none of which is cited by Templer and Arikawa. Full-score equivalence is virtually never found, although weaker forms of equivalence (e.g. factor-loading equivalencies) may be. It is not found even when comparing different age groups within the United States (Hertzog & Bleckley, 2001). Given results like this, the assumption of full score comparability across nations as diverse as Denmark, Nepal, and Guinea is hardly warranted.

We conclude that Templer and Arikawa's conclusions rest on statistical analyses that make more assumptions about the data than the data can support. This conclusion is not confined to Templer and Arikawa's use of the product–moment (Pearson) correlation. It applies to the rank order (Spearman) correlation as well, as failure of full-score comparability could produce a change in the ranks of scores.

3. Objections based on applications of the logic of science

We now shift our argument to the conclusion itself. A good scientific paper reports data that either (a) discriminate between several different hypotheses, or (b) cannot be explained by any of the current theories on a topic. In either of these cases, the findings change our thinking. This contrasts with a good political argument, in which the important thing is to find data that support the proposer's argument.

Templer and Arikawa observe that their data are consistent with the Lynn-Rushton-Jensen hypothesis that both lighter skin and intelligence are evolutionary responses to the challenge of living in high latitudes.

We agree that Templer and Arikawa's data are consistent with the Lynn-Rushton-Jensen argument, providing one accepts the ancillary argument that IQ scores are valid indicators of conceptual intelligence, across different nations and cultures. Even if the ancillary argument is acceptable, the Templer and Arikawa data is consistent with many other explanations. We will return to both these assumptions below. First, though, let us consider the status of the Lynn-Rushton-Jensen argument as a scientific hypothesis.

Obviously, no one can turn the clock back 65,000 years and observe the challenges faced by peoples, of whatever skin color, as they migrated out of Africa and across the globe. On the Eurasian land mass there have been so many migrations that we have no idea whether light skins arose after dark skinned people migrated to Northern climates, as the Lynn-Rushton-Jensen argument requires, or whether light skins arose in the low latitudes, and the light skinned people happened to migrate northward.

The assumption that the challenges of life in higher latitudes has been historically more challenging than life at lower latitudes is simply that, an assumption. There is no way of answering this question by comparing present-day populations, for culturally dependent technologies have spread all over the globe. All we can do is cite examples. We are sure that Lynn, Rushton, and Jensen have theirs. Here are some of ours.

Arguably the greatest single step forward toward our present civilization was the move to agriculture. There have evidently been at least three separate developments of agriculture; the earliest being in the New Guinea highlands, quite close to the Equator, somewhere around 30,000 years ago. Big steps toward civilization were taken in the low latitudes, by the Egyptians, Mesopotamians, and Maya.

The same sorts of arguments can be made if you compare how dark skinned and light skinned peoples reacted to environmental challenges in historic times. The fair-skinned Norse seafarers made amazing voyages in open boats, but so did the darker skinned Polynesians, at about the same time. For that matter, when the Greenland ice cap expanded in the Middle Ages the Norse were unable to adapt, while the darker skinned Inuit (who arrived in Greenland after the Norse) did adapt and survive to this day (Diamond, 2005).

What does all this prove? Nothing, except that the Lynn-Rushton-Jensen argument is simply a "just so" story. It is impossible to prove or disprove. You can make a rhetorical argument for the hypothesis, by selectively citing some evidence and ignoring others, or you can make a rhetorical argument against the hypothesis, by the same mechanisms. Such arguments are perhaps acceptable in law and politics, but not in science.

Suppose, for a moment, that we lay aside our views of the Lynn-Rushton-Jensen hypothesis and the quality of the Templer and Arikawa data. The Templer and Arikawa study is still of very low, if any, scientific value because it does not discriminate between alternative explanations.

Templer and Arikawa themselves agree that the data are consistent with Diamond's argument that the historic flow of ideas East and West across cultures is facilitated by the geography of the Eurasian land mass, while geography inhibits the flow of ideas in the North–South direction. If we apply Diamond's theory to the Templer and Arikawa data, and accept the cross-cultural validity of skin scores, the skin color–IQ correlation is a happenstance. Light skinned people happened, for unknown reasons, to occupy the Eurasian continent. These people were subjected to a flow of ideas that resulted in societies that were intellectually challenging to the individuals in them. This set of events could produce higher intelligence in the people of the challenging societies. Such intelligence might be because of evolutionary pressures, if there was a reproductive advantage to being of high intelligence, or because the challenging society itself fostered intellectual development in its members, for example, by superior schooling or encouraging analytic modes of thought. The latter argument is a historical-geographic analogue to the more recent observation of changes in IQ scores over Western societies in recent times (Flynn, 1987).

A third class of arguments is directed at the validity of the IQ score as an index of conceptual intelligence outside of developed industrial societies. According to this argument light skinned Europeans and North Americans developed certain tests, as indices of the

mental competencies relevant to success in their societies. To the extent that these tests are not relevant to the mental competencies required in societies that, historically, have been developed by darker skinned people, test scores will fall. Therefore, there will be an IQ–skin color correlation, but the reasons for it are not very interesting.

This hypothesis should be taken very seriously. Culture-specific knowledge, including knowledge of problem-solving methods, has important survival value, and thus is far more important to people in that culture than knowledge of the facts and problem solving skills evaluated in Western-developed intelligence tests. For instance, in a study in rural Kenya (Sternberg et al., 2001), children's knowledge of natural herbal medicines used to treat parasitic infections was negatively correlated with scores on conventional tests of intelligence. Similar findings were obtained in a study of the Yup'ik Inuit in Alaska (Grigorenko et al., 2004). The participants in this study did not do at all well on conventional IQ tests, but had the knowledge and skills required for hunting, fishing, and overland travel in an extremely harsh environment.

We are not arguing that the skills required to take conventional IQ tests, including the so-called culture-fair tests, are totally irrelevant to cognitive performance in non-industrialized societies. We are arguing that these tests, which were properly developed to predict performance in industrialized societies, fail to evaluate some cognitive skills that are important in various non-industrialized societies, and probably evaluate some skills that are of marginal relevance in those societies. The extent to which the tests are valid measures of intellectual competence will depend upon the societies involved. Furthermore, very few, if any, cultures on the globe today are untouched by the industrial societies, so relevance of a test will vary over time, as societies change.

In fact, we can see this in historical studies of Western societies. We like to imagine that our tests, especially of fluid abilities, are somehow culture-fair or culture-free. They are not. Not only do the supposedly culture-fair or culture-free tests show greater differences across cultures than do supposedly more loaded tests, they also show a larger cohort effect (Flynn, 1984, 1987; Neisser, 1998). As the cohort effect must be environmental because of the short duration in which it took place, the larger effects for fluid tests can only indicate that they are more, not less affected by experience than the seemingly more culturally loaded tests of crystallized abilities.

This is not an argument that IQ tests are totally irrelevant in developing countries, just that they are relevant to a lesser degree in developing countries than in the countries where the tests were developed. According to this argument the association between IQ score and other indices of personal success should be diminished in non-Western countries. There is a substantial body of evidence showing that this does occur (Sternberg, 2004): Indeed, conceptions of success, as well as those of intelligence, vary across cultures. The types of success that matter so much in the United States, such as money, may simply have less value elsewhere. The academic skills that are so important to Western conceptions of intelligence are not so highly valued elsewhere (Sternberg & Kaufman, 1998).

Any of these classes of arguments could be developed in different ways. Our point is simply that an IQ–skin color correlation could be predicted by theories based on any one of the following world views:

(1) There are common biological (genetic) determinants of skin color and conceptual intelligence, which can be measured by an IQ test.
(2) There is no common biological determinant of skin color and conceptual intelligence, which can be measured by an IQ test, but historical happenstances have produced a correlation between the two.
(3) The IQ test is itself a cultural artifact. Due to historical happenstance, it was developed by light-skinned people, and has validity to the extent that the examinee's culture resembles that of the culture in which the test was developed. The IQ–skin color correlation tells us little or nothing about a possible correlation between skin color and conceptual intelligence.

So why is this correlation a useful piece of information in advancing our science?

4. A concluding comment

The Templer and Arikawa paper is about racial differences in intelligence. It would be naïve to ignore the social and political ramifications of assertions like theirs, and the assertions that may be made about their findings in the secondary literature. Therefore we want to be clear about what we are and are not saying.

We are saying that the Templer and Arikawa article represents dubious research, at the technical level, and should not have been published on those grounds alone.

We are not saying that the topic of biological and genetic determinants of intelligence should not be pursued. This is a separate issue.

We also are not saying that group differences in intelligence should not be studied. There are many different groupings of human peoples. We know that some of these genetic groupings have ramification for biological conditions, such as sickle cell anemia. The behavior genetics data suggest strongly that there are genetic groupings for cognitive skills. Exploring this issue is reasonable science. The social benefits of such science will depend upon how the results are used. This is an issue outside of science itself.

Note, though, that we said "group differences in intelligence" and referred to genetic groupings. Whether it is a useful venture to pursue cross-racial studies of intelligence, as races are defined by the authors, is a rather different issue. Some question whether biological races even exist (Sternberg, Grigorenko, & Kidd, 2005).

Socially, skin color is a major index that is used to assign individuals to groups. Perhaps some future study combining molecular genetics and psychology will show either that the biological mechanisms that produce skin color overlap with those that produce intelligence or that the two mechanisms are distinct but happen, due to historical developments, to be correlated. Or perhaps they will show that the relevant genes are uncorrelated.

Because of the social ramifications, such research should be done, but should be done carefully. People who wrap themselves in the mantle of Galileo, claiming that those who urge caution are trying to hide truth, miss the point. Bad research on this topic should be discouraged much more strongly than bad research on other, less charged topics.

We maintain that the Templer and Arikawa paper is an almost prototypical example of such bad research. To summarize, these are our observations:

(1) Their measure of "skin color" is in fact a measure of social stereotypes about skin color. Social stereotypes and IQ scores may or may not be correlated. If they are the explanation certainly is not biological.
(2) The Lynn and Vanhanen estimates of national IQ are technically inadequate for several reasons, elucidated in the text.
(3) The statistical analyses offered require the assumption of full score equivalence of IQ scores, which cannot be maintained.
(4) The statistical analyses, when properly done to allow for size of the country, are dependent upon having accurate measures of both predominant skin color and IQ in the larger countries, but
(5) The concepts of predominant skin color and average IQ are highly suspect as meaningful concepts for countries of any size and/or social diversity.
(6) Equating intelligence test score (IQ) with intelligence in a conceptual sense, across cultures, is an extremely dubious operation.
(7) The Lynn-Rushton-Jensen hypothesis, which Templer and Arikawa purport to test, is a rhetorical argument rather than a testable scientific hypothesis.
(8) And even if we suspend disbelief about all the above issues, a worldwide correlation between IQ scores and skin color could be explained by many, philosophically contrary hypothesis. Therefore the fact, if it is a fact, is of no scientific value.

We are confident that the majority of regular readers of this journal, who generally have a substantial background in the relevant science, are unlikely to think much of the Templer and Arikawa article. However, the publication of an article in a referred journal carries a certain cachet with it. We are concerned that careless or socially motivated references to the Templer and Arikawa article in the secondary literature will do harm. If the Templer and Arikawa article was a good research, then we would say that the facts are the facts, and that debate about the causes of facts should always be encouraged, however uncomfortable a particular fact may leave us. That is not the case here. Given the blatant inadequacies of the research, we believe that publication of the Templer and Arikawa article was unfortunate.

References

Diamond, J. (1997). *Guns, germs and steel: The fates of human societies*. New York: W.W. Norton.

Diamond, J. (2005). Collapse: How societies choose to fail or succeed. New York: Viking.

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

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

Grigorenko, E. L., Meier, E., Lipka, J., Mohatt, G., Yanez, E., & Sternberg, R. J. (2004). Academic and practical intelligence: A case study of the Yup'ik in Alaska. *Learning and Individual Differences*, *14*, 183–207.

Hertzog, C., & Bleckley, M. K. (2001). Age differences in the structure of intelligence: Influences of information processing speed. *Intelligence*, *29*(3), 191–217.

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

Lynn, R. (1991). Race differences in intelligence: A global perspective. *Mankind Quarterly*, *31*, 255–296.
Lynn, R., & Vanhanen, T. (2002). *IQ and the wealth of nations*. Westport, CT: Praeger Publishers/Greenwood Publishing Group, Inc.
Neisser, U. (Ed.). (1998). *The rising curve* (pp. 95–112) (pp. 95–112). Washington, DC: American Psychological Association.
Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics*, *6*(1), 101–115.
Rushton, J. P. (1995). *Race, evolution, and man*. Princeton: Princeton University Press.
Sternberg, R. J. (2004). Culture and intelligence. *American Psychologist*, *59*(5), 325–338.
Sternberg, R. J., Grigorenko, E. L., & Kidd, K. K. (2005). Intelligence, race, and genetics. *American Psychologist*, *60*(1), 46–59.
Sternberg, R. J., & Kaufman, J. C. (1998). Human abilities. *Annual Review of Psychology*, *49*, 479–502.
Sternberg, R. J., Nokes, K., Geissler, P. W., Prince, R., Okatcha, F., Bundy, D. A., et al. (2001). The relationship between academic and practical intelligence: A case study in Kenya. *Intelligence*, *29*, 401–418.
Van de Vijver, F., & Leung, K. (1997). Methods and data analysis of comparative research. In John W. Berry, Ype H. Poortinga, & Janak Pandey (Eds.), *Handbook of cross-cultural psychology, Vol. 1: Theory and method (2nd ed.)* (pp. 257–300). Needham Heights, MA: Allyn & Bacon.

Earl Hunt
University of Washington, United States
E-mail address: ehunt@u.washington.edu.
Corresponding author.

Robert J. Sternberg
Yale University, United States
E-mail address: robert.sternberg@yale.edu.

doi:10.1016/j.intell.2005.04.004

The Jensen and the Hunt and Sternberg comments: From penetrating to absurd

Abstract

We praised the comments of Jensen and regard most of the contentions of Hunt and Sternberg as absurd. It is ridiculous to question the validity of the skin color map and its application since meaningful group differences and meaningful correlations between temperature and skin color were found. It was inappropriate for Hunt and Sternberg to attribute prejudices and erroneous preconceptions to our raters who were assigned a task that inherently permits very minimal subjective interpretation. The suggestion of Hunt and Sternberg that higher intelligence evolves in equatorial people is incongruent with the correlation or 0.62 between cranial capacity and distance from the equator reported by Beals et al. Hunt and Sternberg failed to provide a balanced perspective in their critique of the Lynn and Vanhanen international presentation of IQs.

1. Comments on the Jensen commentary

Jensen provided an outstanding commentary. He politely and articulately understated his case when he said "The collection of mean IQs in the various countries (whether *N*-weighted or not) almost certainly fall short of the degree of reliability and validity attainable with psychometric tests administered under laboratory conditions." He went on to say that the data have sufficient overall precision for "ecological correlations" that are widely accepted in epidemiological medical research. He maintained that the aggregation of test scores "average out" the unique and irrelevant sources of variance among individuals.

Jensen pointed out that our correlations between skin color and IQ on the international level are consistent with his review of 18 published studies with African- Americans that yielded an average correlation of about 0.20. We recognize that a correlation of 0.20 accounts for only 4% of the variance. It is, however, understandable that the correlation is much lower than in our international study because the skin color variance is obviously smaller.

Jensen suggested that we may have been too conservative in dismissing the possibility of cause and effect relationship between skin color and IQ because of the genetic phenomenon of pleiotropy in which a single gene has more than one phenotypically quite different effect. Jensen brilliantly suggested research using sibling pairs that tests the hypothesis that the lighter skinned sibling tends to have higher intelligence. We believe that such a study would be feasible. This could be carried out by finding a school that has both student pictures and student IQs or achievement or aptitude tests that correlate rather highly with IQ such as the SAT. Another research idea we propose is that of correlating skin color with head size in fraternal twins. The *N* would have to be very large for adequate power because the correlation of IQ with external head measurements is lower than with MRI determined brain size.

2. Comments on the Hunt and Sternberg commentary

In contrast to Rushton and Lynn who maintain that higher intelligence evolves in colder climates, Hunt and Sternberg suggested that the challenge of fending off a myriad of parasitic diseases should make equatorial people more intelligent. Their idea is an intriguing one. The research evidence, however, does not support their contention. In our article under discussion temperature correlates negatively and not positively with IQ. An even more telling blow to the Hunt and Sternberg argument is the correlation of 0.62 between cranial capacity and distance from the equator reported by Beals et al.

Hunt and Sternberg placed much more emphasis on the skin color correlations than the temperature correlations. This obscures the fact that the Templer and Arikawa study was primarily designed to test the contention of Rushton and Lynn that people in colder climates tend to have higher intelligence. Skin color and IQ were the two principal variables in the design of the study. The negative correlations between temperature and IQ strongly support the position of Rushton and Lynn. Skin color was conceptualized as a climatic variable, a multigenerational reflection of the climate one's ancestors have lived in for thousands of years. The very high correlation of IQ and skin color provides "frosting on the cake."

Hunt and Sternberg acknowledged that the high skin color correlations between ratings of the three graduate students demonstrate reliability. They asserted, however, that the raters "may share the same implicit theories, prejudices, erroneous perceptions, or whatever. For example the fact that the three judges from the Salem witch trials may have shared the same views as to which of the accused were witches did not make the accused witches." We regard this argument as absurd. One of the raters never met or talked to the other two raters and is doing her graduate work at a different school over a thousand miles away. More importantly, the nature of the task contains an extremely small subjective element. The raters used only two things—the skin color map of the world that does not have delineated national boundaries and maps of their choice that do show national boundaries. The majority of the countries were covered by a shading representing only one skin color. In retrospect, we really did not have to have the raters consider those countries. No element of judgement was involved with these countries. The concern of Hunt and Sternberg about India has some justification. In fact, we had the same concern because it has more skin colors than any other country. Because of this concern we computed the correlation between skin color and IQ without India. The correlations for the 128 and 129 countries were remarkably similar and both round to -0.92. This was reported to the Editor in our extensive communication.

We dismiss the Hunt and Sternberg criticism of the raters not having visited the countries and having no relevant information about these countries (even if it is correct) as irrelevant. In fact, having visited a country may even produce more biases than it corrects. If a visitor from another continent should go to Vancouver in January he or she may infer that Canada has mild winters. One of the authors, Arikawa, was surprised to find out that her native country of Japan had a mean IQ of only 105 after having lived in the country for a quarter of a century.

Students in undergraduate test and measurements classes are taught that correlations and group differences support the validity of measurement instruments. Our negative correlations between temperature and IQ support the validity of the skin color map. In regard to group differences, there is no skin color shading overlap between the European countries and the Black African countries. Do Hunt and Sternberg doubt that temperature and skin color are correlated? Do they doubt that Black Africans tend to be darker than Europeans?

The skin color map is consistent with common knowledge. Southern Europeans tend to be darker than northern Europeans. Africans below the Sahara desert tend to be darker than Africans (e.g. Arabs and Berbers) north of the Sahara desert. A reasonable assessment is that the skin color map may have limitations. It would be preposterous to say that it has little or no validity.

Hunt and Sternberg argued that one reason our skin color–temperature correlations are spurious is that non-Western people were tested with Western-developed measures of intelligence. Their argument should not be categorically rejected. In fact, we said "Although both of these instruments would appear to be void of educational and specific culture content, it cannot be assumed that they are equally effective in measuring intelligence around the world." Relevant to this issue is a Templer and Arikawa study now in its final stages. We computed the correlations between IQ and skin color in three continents. The correlations are -0.86 for Africa, -0.55 for Asia, and -0.63 for Europe. It would be difficult to argue that the Ravens Progressive Matrices and the Cattell Culture Fair Test are inappropriate for Europeans.

The present authors agree with Jensen and agree with Hunt and Sternberg that the methodology of the Lynn and Vanhanen international aggregation of IQs is far from perfect. The standardizations of the Wechsler tests of intelligence are also not perfect. There is, however, no other international aggregation of IQs. It is most unlikely that the authors of the individual IQ studies consistently tested below national average participants in the warmer counties and consistently tested above national average participants in the colder countries. In general, imperfections in measurement instruments are more likely to attentuate than inflate correlations. A correlation of 0.92 between two worthless instruments is not possible. On the other hand, we urge that this correlation not be viewed as immutable.

Hunt and Sternberg failed to provide a balanced perspective in discussing the methodology limitations of Lynn and Vanhanen. They criticized these authors for averaging the means of IQs that have Ns of different sizes. This criticism would possibly be appropriate if Site A in a given country had a population of 1000 and 100 were tested and Site B had a population of 500 and 50 were tested. If, however, the selection of participants approached being haphazard as Hunt and Sternberg implied, it would not be mathematically justifiable to apply different weights to different sites. Furthermore, Lynn and Vanhanen reported high reliability for their national IQs using the 45 countries that have two measures of IQ and the 15 that have three or more (in which case the two extreme mean IQs were employed). The correlation between the two IQs in the 60 countries was 0.94. In regard to the use of IQ tests in non-Western populations, Lynn and Vanhanen pointed out that people from a variety of cultures, including Ugandans and Black South Africans, have the same pattern of intercorrelations and an identifiable *g* factor on ability tests. Lynn and Vanhanen provided evidence that reaction time, which correlates positively and substantially with IQ, has the same rank order as IQ in five countries of the world. Hong Kong, Japan, Britain, Ireland and South Africa rank in descending order in both IQ and six different reaction time measures. The correlations are 0.94, 0.89, 0.96, 0.83, 0.73 and 0.85. Lynn and Vanhanen reminded their readers that East Asians tend to score higher on Western developed IQ tests than do Europeans and Americans.

In regard to Hunt and Sternberg saying that the publication of our article is unfortunate, we feel the need to point out that the Editor did not quickly or impulsively accept it. It was refereed by three psychologists who are recognized as leading authorities in the world in the area of intelligence. The Editor and the authors engaged in extensive communications over a period of months. Further input of the referees was obtained. Major concessions were made to the extremely critical reviewer. Three revisions were required by the Editor, and the Editor finally accepted with the understanding that comments to our article will also be published.

Donald I. Templer
Alliant Internationial University, United States
E-mail address: donaldtempler@sbcglobal.net.

Hiroko Arikawa
Springfield, MO, United States

doi:10.1016/j.intell.2005.04.005