

- 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.
 © 2005 Elsevier Inc. All rights reserved.

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 large 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/Greenwood.

- 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 of 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.

© 2005 Elsevier Inc. All rights reserved.

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.