Psychology, Public Policy, and Law
2005, Vol. 11, No. 2, 328–336

1076-8971/05/$12.00 DOI: 10.1037/1076-8971.11.2.328

WANTED: MORE RACE REALISM, LESS MORALISTIC FALLACY

J. Philippe Rushton
The University of Western Ontario

Arthur R. Jensen
University of California, Berkeley

Despite repeated claims to the contrary, there has been no narrowing of the 15- to 18-point average IQ difference between Blacks and Whites (1.1 standard deviations); the differences are as large today as they were when first measured nearly 100 years ago. They, and the concomitant difference in standard of living, level of education, and related phenomena, lie in factors that are largely heritable, not cultural. The IQ differences are attributable to differences in brain size more than to racism, stereotype threat, item selection on tests, and all the other suggestions given by the commentators. It is time to meet reality. It is time to stop committing the "moralistic fallacy" that good science must conform to approved outcomes.

In our target article (Rushton & Jensen, 2005), we proposed a hereditarian model—50% genetic–50% environmental—to explain the 15- to 18-point average IQ difference (1.1 standard deviations) between Blacks and Whites. We reviewed the worldwide distribution of test scores, the *g* factor of mental ability, the heritability of within- and between-groups differences, the relation of brain size to intelligence and of race differences in brain size, regression to the mean, cross-racial adoption studies, racial admixture studies, and data from life-history traits and human origins research. We were unable to identify (in Section 12 of Rushton & Jensen, 2005) any reliable environmental contribution to the Black–White IQ difference, including the non-*g* Flynn effect (i.e., the secular rise in IQ scores). We also found that on many dimensions, East Asian–White differences were a mirror image of Black–White differences. In Section 14, we concluded in favor of an even stronger hereditarian model—80% genetic–20% environmental—based on Jensen's (1998, p. 443) "default hypothesis" that, by adulthood, genetic and environmental factors carry the same weight in causing group differences as they do in causing individual differences.

Gottfredson (2005) is the only commentator who confronted head-on all the empirical, theoretical, and moral issues. The other commentators (Nisbett, 2005; Sternberg, 2005; Suzuki & Aronson, 2005) sidestepped the totality of the three-way race–behavior matrix shown in our Table 3. They invoked one or other of the culture-only refrains, that "race" is only "skin deep"; if not, then any difference is too small to matter; if not, then it is due to cultural factors such as statistical artifacts, insensitive tests, racism, stereotype threat, and poverty; if not, then it is poor form to talk about it. They also offered the usual culture-only promissory notes that the Black–White IQ gap can be reduced by economic improvements, interventionist programs, culture-friendly assessment systems, and nonweighted

J. Philippe Rushton, Department of Psychology, The University of Western Ontario, London, Ontario, Canada; Arthur R. Jensen, School of Education, University of California, Berkeley.

Correspondence concerning this article should be addressed to J. Philippe Rushton, Department of Psychology, The University of Western Ontario, London, Ontario N6A 5C2, Canada. E-mail: rushton@uwo.ca

models of gene–environment interaction. Their examples only confirm what we described in Sections 2, 13, and 14: Culture-only theory is a degenerating research paradigm.

Are Black–White IQ Differences Narrowing?

Nisbett (2005) provided the most empirically forceful of the rebuttals. He claimed that the Black–White IQ difference had decreased to only 10 points in magnitude (<0.70 standard deviations) and that it could be eliminated altogether within 20 to 60 years. He based this assertion on a purported narrowing of the Black–White difference on school achievement tests (reading, vocabulary, and mathematics), which he then extrapolated to the IQ differences.

Reality, however, is stubborn. Jensen (1998, pp. 375–376, n. 33, 407–408, 494–495) showed that gains in scholastic achievement do not equal gains in *g,* and the Black–White differences in *g* are as large as ever, even for measures of reaction time. Jensen's conclusion dovetails with a meta-analysis by Roth, Bevier, Bobko, Switzer, and Tyler (2001) that we cited at the opening of our target article. They found a mean effect size of 1.1 standard deviations that ranged from 0.38 to 1.46 (based on a sample of 6,246,729 from corporate, military, and higher education samples) depending on the *g* loading of the test. On the question of whether the difference was diminishing, they suggested any reduction was "either small, potentially a function of sampling error . . . or *nonexistent for highly g loaded instruments* [italics added]" (Roth et al., 2001, p. 323).

In her commentary, Gottfredson (2005) underscored this message with evidence that no narrowing had taken place in average Black–White differences. She contrasted Black–White differences on highly *g*-loaded "IQ tests" with those on less *g*-loaded "school achievement tests." Gottfredson found that Black–White differences on IQ tests remained constant at 1.0 standard deviation throughout the 20th century. She agreed that the differences on school achievement tests did narrow slightly from 1.07 to 0.89 standard deviations from the 1970s to the 1990s when the National Assessment of Educational Progress collected data on 9- to 17-year-olds. However, as she then pointed out, even this 20% reduction in educational achievement (a) had occurred by the mid-1980s and no longer continues, (b) is compatible with the group differences in *g*, and (c) does not contradict the hereditarian hypothesis.

These variable Black–White differences are explained by Spearman's (1927) hypothesis, which states that Black–White IQ differences are "most marked in just those [tests] which are known to be saturated with *g*" (p. 379; see Section 4 of Rushton & Jensen, 2005). The differences are lower on specific tests of memory, or arithmetic and spelling, than they are on general reasoning and transforming information. One implication is that test constructors could in principle reduce the Black–White difference to zero (or even reverse it) by including only non-*g* items (or those negatively loaded on *g*). However, they would then be left with a test that had little or no predictive power. Roth et al.'s (2001) meta-analysis concluded: "Overall, the results for both industrial and educational samples provide support for Spearman's hypothesis. That is, black–white differences on measures of cognitive ability tended to increase with the saturation of *g* in the measure of ability" (Roth et al., 2001, p. 317).

There is in fact no good evidence, contrary to Nisbett (2005; and Suzuki & Aronson, 2005), that *g* is malleable by nonbiological variables. That would require not just evidence that training produces higher scores but evidence of broad transfer of training effects to other highly *g*-loaded tasks. Extrapolation of the trends into the future may be like extrapolating the non-*g* secular rise in IQ scores (the Flynn effect; see Section 12). That the Flynn effect is not a Jensen effect (i.e., did not have a loading on the *g* factor) was corroborated by Wicherts et al. (2004). This is consistent with the lack of convergence of White and Black means across decades despite the overall rise in IQs.

Two recent monographs show just how wide the achievement gap between Blacks and Whites remains. First, Thernstrom and Thernstrom (2003) comprehensively documented the scale of the Black deficiency: For example, in reading, history, geography, and mathematics, 12th-grade Black students do not do as well as eighth-grade White students. The authors showed, moreover, that despite numerous, often well-publicized, countywide projects (such as the $2 billion program in affluent Montgomery County, Maryland, as well as the Kansas City, Missouri, school district, under judicial supervision since 1985), no plan has yet made a replicable dent in the Black–White achievement gap (despite low student–teacher ratios and computers in every classroom). Second, Ogbu (2003) studied the persistent underachievement of Black children in the well-to-do suburb of Shaker Heights, Ohio, as a result of concern raised by their (Black) parents, often highly paid professionals who had moved to the area specifically for its schools. The Black students did better than Black students elsewhere, but there were huge gaps between the Blacks and their non-Black counterparts. Instead of genetic differences in intelligence, both books offer variations on the usual culture-only explanations: poor schools, prejudice, stereotyping, low expectations, and alienation from White cultural domination. Nor do they consider regression to the mean (Section 9) or other genetically influenced traits that differentiate the races and affect attitudes to schoolwork (Section 10).

Racial Admixture Studies: Direct Versus Indirect Evidence of Heritability

Nisbett (2005) cited seven empirical studies on people of mixed race (based on self-reported ancestry, skin color, and blood groups) as "direct evidence" for the "nil" heritability of Black–White differences. He claimed these outweighed those we had presented (in Rushton & Jensen, 2005, Sections 7 and 8). It should be noted that Nisbett's studies are peculiarly old, the mean year of publication being 1960 (median year 1966; range = 1930 to 1977). Most are actually very weak and nondecisive, not having been replicated even once. Some are so old and recycled that Jensen (1973; see also 1998, pp. 478–483, 612) dealt with them 30 years ago! The blood-group studies could be repeated with better sampling and methods of analysis, but probably never will be because a more powerful tool, DNA analysis, is now available for this purpose. In Section 8, we discussed the DNA methods that can ascertain degree of White ancestry in Black populations. Many other DNA markers identify Black–White differences regardless of how divergent the African ancestry. They have been recommended for evaluating

admixture in genetic studies of disease (Collins-Schramm et al., 2002), and we recommend them for genetic studies of IQ.

More generally, we do not share Nisbett's contention that "direct" evidence is more relevant than "indirect" evidence unless, of course, the quality, quantity, and consistency of the direct evidence are also stronger than the indirect evidence. Much of evolutionary theory, genetics, chemistry, and physics are essentially based on what Nisbett would call indirect evidence. The hereditarian model of an 80% genetic–20% environmental weighting for the Black–White IQ difference is based on the hypothetico-deductive method (Sections 2 and 14), not a patchwork of narrow, often inconsistent or unreplicated facts. Our "indirect" evidence includes the fact that (a) the gene–environment architectural matrix is the same for both races (Section 5); (b) inbreeding scores from Japan predict mean Black–White differences in the United States (Section 5); (c) regression to the mean operates consistently in both races (Section 9); (d) psychometric *g* is one and the same factor in both Whites and Blacks (Section 4); and (e) race differences are greatest on the *g* factor extracted from both IQ tests and reaction time tasks (Section 4).

How do the critics explain the fact that the Black–White difference is greater on backward than on forward digit span memory, or on the more complex rather than simple reaction time measures—exactly as predicted by Spearman's (1927) hypothesis? How do they explain the fact that Black students from families with incomes of $80,000 to $100,000 score considerably lower on the SAT than White students from families with $20,000 to $30,000 incomes? How do they explain why social class factors, all taken together, only cut the Black–White achievement gap by a third? Culture-only theory cannot predict these facts; often its predictions are opposite to the empirical results.

African IQ Scores

Sternberg's (2005) and Suzuki and Aronson's (2005) commentaries about IQ studies from sub-Saharan Africa are written as though we are not aware that African children suffer from parasitic illnesses and malnutrition, speak languages other than English, grow up in cultures of violence, or that mediated learning interventions show increases in African IQ scores. We cited three studies on mediated learning (Section 3), including the one by Skuy et al. (2002) that Suzuki and Aronson referred to in detail, on which Rushton was a coauthor!

Rushton's series of studies in South Africa (see Rushton & Jensen, 2005, Section 3) sought to examine further the well-replicated reports of an African population mean of IQ = 70. He tested to see if IQ scores from highly select students at the prestigious University of the Witwatersrand in Johannesburg were consistent with the mean IQ of 70 reported for the general African population (Lynn & Vanhanen, 2002). The results from seven studies conducted at universities in South Africa, including those by other investigators, yield a median IQ of 84 (range = 77 to 103). Assuming that African university students are 1 standard deviation (15 IQ points) above the population mean, the finding of a median IQ of 84 corroborates the general population mean of 70. Although Rushton's mediated learning study with Skuy et al. (2002) on first-year psychology students did raise the IQ of the African students from 83 to 97, this is still low for students

at a leading university. Moreover, as we mentioned in Section 3, evidence shows that "coaching" or "teaching-to-the-test" has the effect of denuding the test of its *g* loading (te Nijenhuis, Voskuijl, & Schijve, 2001).

There can be little doubt about the replicability of the mean African IQ of 70, or the impartiality of the investigators, for studies continue to report low scores. In Kenya, Sternberg et al. (2001; see also Sternberg, 2005) administered the Colored Progressive Matrices to 85 children ages 12 to 15 years who scored 23.5 out of 36, an IQ equivalent of about 70. In Tanzania, Sternberg et al. (2002; also Sternberg, 2005) gave the Wisconsin Card Sorting Task to 358 children ages 11 to 13 who received a perseverative error score of 18.53. Although procedural differences may make the normative comparison problematic, as it stands, this score is equivalent to the fifth percentile on American norms for 12-year-olds (IQ = 75). After training on how to sort attributes, the children's scores went up to 16.5 (lower scores meant fewer errors), but this was still only at the ninth percentile on American norms (IQ $<$ 80).

We accept as nonarguable that intervention strategies in Africa such as the elimination of tapeworms, improved nutrition, and provision of electricity, schools, and hospitals will raise test scores. However, we predict they will not remove the substantial differences in average IQ between Africans and Europeans, and that African Americans and other mixed-race populations will continue to average between these "pure" types because of White admixture. As regards Suzuki and Aronson's (2005) reference to "a context of racism and colonialism that, in turn, creates and shapes stereotypes" (p. 324), it should be noted that many of the African countries showing a mean IQ = 70, such as Nigeria and Ghana, have been independent for half a century (and the Caribbean Island of Haiti for one and a half centuries), with no documented improvement in cultural achievement or IQ scores.

Around the world, mean IQs differ much less *within* major population groups than *between* them. Whites have IQs close to 100 whether they live in Europe, Canada, Australia, New Zealand, or South Africa, whereas Blacks in sub-Saharan Africa have IQs closer to 70 regardless of whether they live in East, West, Central, or Southern Africa—or whether the data were collected in the 1920s or the 2000s (Lynn & Vanhanen, 2002). The IQ of Blacks in the United States is around 85 and hence substantially higher than the IQs of Blacks in sub-Saharan Africa. There are two explanations for this. The first is that American Blacks have about 25% White ancestry. According to genetic theory this would raise their IQs above the level of Blacks in Africa. The second is that American Blacks enjoy much higher standards of living, nutrition, education, and health care than they have in societies run by Blacks. Living in a White society has raised rather than lowered the IQs of American Blacks.

Genetic factors explain the worldwide pattern in a way that culture-only theory has not. The worldwide pattern contradicts the hypothesis that the low IQ of American Blacks is due to "White racism." For instance, Mackintosh (1998) wrote, "it is precisely the experience of being black in a society permeated by white racism that is responsible for lowering black children's IQ scores" (p. 152). The IQs of Blacks in Africa is compelling evidence against this theory. The theory that White racism has been responsible for the low IQ of American Blacks always had an ad hoc quality to it because "racism" has had no adverse impact on the

intelligence of East Asians and Jews, who average higher scores than do Europeans (Section 1).

Brain-Size Differences

Brain size and its relation to intelligence are crucial for an evolutionary understanding of the origin of race differences in behavior. Both magnetic resonance imaging (MRI) and external head size measures show that brain size is related to IQ within race. Moreover, the three-way pattern of East Asian–White–Black differences in brain size that is found in adulthood (1,364 cm^3, 1,356 cm^3, and 1,267 cm^3, respectively; see Rushton & Jensen, 2005, Section 6) is detectable at birth. The findings on race and brain size have been repeatedly replicated and found to be robust across variations in measures, methods, and subject samples. How do our critics handle this evidence? Rather than refuting or challenging this evidence, our critics completely ignore it.

If two groups differ by 1 standard deviation in brain size and the correlation between brain size and IQ is 0.40, then they will differ by 6 IQ points! Sarich and Miele (2003) estimated the Black–White difference in brain size as 0.8 standard deviations, hence a 5-point IQ difference is attributable to brain size alone. When purer measures of *g* are used (Jensen, 1998) or a larger standard deviation for brain size (Rushton, 2000), the regression of brain size on *g* comes to over half the *g* difference.

The Moralistic Fallacy and Public Policy

The naturalistic fallacy, identified by philosopher David Hume (1711–1776), occurs when reasoning jumps from statements about what *is* to prescription about what *ought to be.* An example of the naturalistic fallacy would be to support warfare if scientific evidence showed that it was to some degree part of human nature. (Warfare may or may not be supportable; the point is only that it is not *logical* to derive "ought" from "is.") The converse of the naturalistic fallacy is the moralistic fallacy, which occurs when reasoning jumps from prescriptions about what *ought to be* to statements about what *is.* It was coined by Harvard University microbiologist Bernard Davis (1978) as a response to calls for ethical guidelines for studying what could purportedly become "dangerous knowledge," such as the genetic basis of IQ. Davis reasoned that chilling an area of inquiry on moral grounds fixes our knowledge in that area, so it becomes, in effect, an illogical effort to derive an "is" from an "ought." An example of the moralistic fallacy is to claim that because warfare is wrong, it cannot be part of human nature.

One corollary of the moralistic fallacy is the demonizing of those who refuse to observe it. Another is that someone must be blamed whenever Nature stubbornly refuses to conform. Because Blacks and Whites *ought* to be equal in IQ and educational outcome but still are not, some who adopt a moralistic position hold, in effect, that White people's attitudes are largely to blame (e.g., Ogbu, 2003; Thernstrom & Thernstrom, 2003). Both fallacies are conjoined when it is argued that whereas minority dislike of Whites is "natural" (because of mistreatment, or because of feeling "culturally dominated"), White prejudice is inherently bigoted and "unnatural."

Sternberg (2005) questioned whether we showed "good taste" (p. 300) in

researching the hereditarian hypothesis in place of culture-only alternatives such as poverty and racism, and he asked, "What good is research of the kind done by Rushton and Jensen supposed to achieve?" (p. 296). This is worth discussing if only because we will never make progress in race relations if we operate on the belief that one segment of society is responsible for the plight of another segment and that belief is false (see also Gottfredson, 2005).

Ever since Gunnar Myrdal's (1944) *An American Dilemma* was cited in footnote 11 of the U.S. Supreme Court's 1954 decision *Brown v. Board of Education of Topeka* (which outlawed racial segregation in the schools), it has become prevalent to attribute the underachievement of Black people to prejudice and discrimination by White people. Myrdal's "Theory of the Vicious Circle" stated: "White prejudice and discrimination keep the Negro low in standards of living, health, education, manners and morals. This, in turn, gives support to white prejudice. White prejudice and Negro standards thus mutually 'cause' each other" (Myrdal, 1944, p. 75). Myrdal rejected the idea that heredity had anything to do with "low Negro standards," instead praising anthropologist Franz Boas for subverting the up-to-then accepted hereditarian perspective.

Myrdal's (1944) tome (1,500 pages comprising 50 chapters and appendices) identified White people's "attitudes" as the main cause of Black people's problems. He contended, "the scientific *facts* of race and racial characteristics of the Negro people are only of secondary and indirect importance . . . *the beliefs held by white people . . . are of primary importance*" (Myrdal, 1944, p. 110, emphasis in original). Although Myrdal himself acknowledged the facts that Blacks averaged a "head slightly longer and narrower; cranial capacity slightly less; . . . pelvis narrower and smaller" (Myrdal, 1944, p. 139), he worried that these findings would lead Whites to conclude that Blacks had "lower reasoning power," which would be an "incorrect interpretation" because "no connection has been proved between cranial capacity and mental capacity" (Myrdal, 1944, p. 140). He also alleged there had been "exposés" of the "distorted . . . measurements" (Myrdal, 1944, p. 91) of racial differences in brain size (cf. Jensen, 1998; Rushton, 2000).

Conclusion

Discussing the totality of the evidence with those who, for whatever reason, refuse to adopt the behavioral genetic or evolutionary perspective, at least when it comes to the nexus of race, intelligence, and genetics, is little more than arguing past each other. There is not space to respond in detail with the data and analyses that refute each and every criticism raised by the commentators. For more information on the *g* factor as the largest common factor in any battery of diverse cognitive tests, see Jensen (1998, chap. 4); on the scientific definition of race, see Sarich and Miele (2003, chap. 8); on whether the Flynn effect is a Jensen effect, see Wicherts et al. (2004); on transracial adoption studies, see Jensen (1998, pp. 472–478); on Ogbu's class-as-caste hypothesis, see Jensen (1998, pp. 511–513); and on stereotype threat, which is a type of test anxiety, see Jensen (1998, pp. 513–515; see also Sackett, Schmitt, Kabin, & Ellingson, 2001). We reviewed all of the relevant evidence on Black–White IQ differences and concluded that hereditarian models of from 50% genetic–50% environmental (Section 2) to 80%

genetic–20% environmental (Section 14) provide a far better fit than the culture-only model of 0% genetic–100% environment.

Expanding on the application of his "default hypothesis" that Black–White differences are based on aggregated individual differences, themselves based on both genetic and environmental contributions, Jensen (2003) proposed "two laws of individual differences": (a) Individual differences in learning and performance *increase* as task complexity increases, and (b) individual differences in performance *increase* with practice and experience (unless there is a low ceiling on proficiency). Consequently, the more we remove environmental barriers and improve everybody's intellectual performance, the greater will be the relative influence of genetic factors (because the environmental variance is being removed). However, this means that equal opportunity will result in unequal outcomes, *within* families, *between* families, and *between* population groups. The fact that we have learned to live with the first, and to a lesser degree the second, offers some hope we can learn to do so for the third.

References

Brown v. Board of Education of Topeka, 347 U.S. 483 (1954).

Collins-Schramm, H. E., Kittles, R. A., Operario, D. J., Weber, J. L., Criswell, L. A., Cooper, R. S., & Seldin, M. F. (2002). Markers that discriminate between European and African ancestry show limited variation within Africa. *Human Genetics, 111,* 566–569.

Davis, B. (1978, March 30). The moralistic fallacy. *Nature, 272,* 390.

Gottfredson, L. S. (2005). What if the hereditarian hypothesis is true? *Psychology, Public Policy, and Law, 11,* 311–319.

Jensen, A. R. (1973). *Educability and group differences*. London: Methuen.

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

Jensen, A. R. (2003, August 23). *Individual differences and population differences in Spearman's g*. Acceptance speech at Kistler Prize Ceremony, Foundation for the Future, Seattle, WA.

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

Mackintosh, N. (1998). *IQ and human intelligence*. Oxford, England: Oxford University Press.

Myrdal, G. (1944). *An American dilemma*. New York: Harper & Row.

Nisbett, R. E. (2005). Heredity, environment, and race differences in IQ: A commentary on Rushton and Jensen (2005). *Psychology, Public Policy, and Law, 11,* 302–310.

Ogbu, J. U. (2003). *Black American students in an affluent suburb: A study of academic disengagement*. Mahwah, NJ: Erlbaum.

Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., III, & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology, 54,* 297–330.

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

Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law, 11,* 235–294.

Sackett, P. R., Schmitt, N., Kabin, M. B., & Ellingson, J. E. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist, 56,* 302–318.

Sarich, V., & Miele, F. (2003). *Race: The reality of human differences*. Boulder, CO: Westview Press.

Skuy, M., Gewer, A., Osrin, Y., Khunou, D., Fridjhon, P., & Rushton, J. P. (2002). Effects of mediated learning experience on Raven's Matrices scores of African and non-African university students in South Africa. *Intelligence, 30,* 221–232.

Spearman, C. (1927). *The abilities of man: Their nature and measurement*. New York: Macmillan.

Sternberg, R. J. (2005). There are no public-policy implications: A reply to Rushton and Jensen (2005). *Psychology, Public Policy, and Law, 11,* 295–301.

Sternberg, R. J., Grigorenko, E. L., Ngrosho, D., Tantufuye, E., Mbise, A., Nokes, C., et al. (2002). Assessing intellectual potential in rural Tanzanian school children. *Intelligence, 30,* 141–162.

Sternberg, R. J., Nokes, C., Geissler, P. W., Prince, R., Okatcha, F., Bundy, D. A., & Grigorenko, E. L. (2001). The relationship between academic and practical intelligence: A case study in Kenya. *Intelligence, 29,* 401–418.

Suzuki, L., & Aronson, J. (2005). The cultural malleability of intelligence and its impact on the racial/ethnic hierarchy. *Psychology, Public Policy, and Law, 11,* 320–327.

te Nijenhuis, J., Voskuijl, O. F., & Schijve, N. B. (2001). Practice and coaching on IQ tests: Quite a lot of *g*. *International Journal of Selection and Assessment, 9,* 302–308.

Thernstrom, A., & Thernstrom, S. (2003). *No excuses: Closing the racial gap in learning*. New York: Simon & Schuster.

Wicherts, J. M., Dolan, C. V., Hessen, D. J., Oosterveld, P. G., van Baal, G. C. M., Boomsma, D. I., & Span, M. M. (2004). Are intelligence test measurements invariant over time? Investigating the nature of the Flynn effect. *Intelligence, 32,* 509–537.