

THE CULTURAL MALLEABILITY OF INTELLIGENCE AND ITS IMPACT ON THE RACIAL/ETHNIC HIERARCHY

Lisa Suzuki and Joshua Aronson
New York University

This commentary highlights previous literature focusing on cultural and environmental explanations for the racial/ethnic group hierarchy of intelligence. Assumptions underlying definitions of intelligence, heritability/genetics, culture, and race are noted. Historical, contextual, and testing issues are clarified. Specific attention is given to studies supporting stereotype threat, effects of mediated learning experiences, and relative functionalism. Current test development practices are critiqued with respect to methods of validation and item development. Implications of the genetic vs. culture-only arguments are discussed with respect to the malleability of IQ.

Rushton and Jensen (2005) review decades of literature to support a genetic basis for the racial/ethnic group hierarchy in intelligence, a position they have held unwaveringly for over 30 years. Their report gives little mention to findings that point to the impact of environment and race (i.e., race as a social construction) on intellectual development or performance—what they term the *culture-only* perspective. We are not among the culture-only adherents as characterized by Rushton and Jensen. While acknowledging the impact of biological factors on intelligence test performance, we have examined the impact of cultural/environmental factors that affect performance on aptitude and achievement measures. Our work, and that of others (e.g., Aronson, 2002; Sternberg, 1996), show us that intellectual performance is much more fragile and malleable than what is often noted in the current literature. The goals of our commentary are to highlight, briefly, assumptions underlying definitions (i.e., intelligence, heritability, genetics, culture, race) and clarify historical, contextual, and testing issues that were only briefly mentioned by Rushton and Jensen. Finally, we comment on the heuristic value and on policy implications of the research.

Problematic Assumptions Underlying Definitions

Rushton and Jensen's (2005) argument rests on particular definitions of intelligence, genetics (i.e., heritability), culture, and race. It is, in part, differences in definitional assumptions that have allowed researchers to claim support for distinctly different perspectives (i.e., environment vs. genetics) based on the same data (Hayman, 1998).

Lisa Suzuki and Joshua Aronson, Department of Applied Psychology, New York University. We would like to thank John Kugler, Leo Wilton, Jacqueline Mattis, and Muninder Ahluwalia for their feedback on earlier versions of this commentary.

Correspondence concerning this article should be addressed to Lisa Suzuki, Department of Applied Psychology, New York University, East Building 239 Greene Street 409, New York, NY 10003. E-mail: las1@nyu.edu

Intelligence

Numerous theories of intelligence have been framed and reframed over the years as scholars have ruminated about what constitutes “intelligence.” In this section we highlight a few of the issues that complicate the linkage between race and IQ as presented by Rushton and Jensen (2005). As Fagan and Holland (2002) argued, IQ scores represent a composite of how well one does in comparison with one’s peers. Test performance is a measure of a person’s intellectual ability that is dependent on one’s genetic makeup and affected by environment and cultural experiences (e.g., informal learning and schooling).

Psychometric definition of g. Although different theories of intelligence have been noted throughout the literature, intelligence has been defined to a large extent by the tests designed to measure it. It should be noted that others have challenged the emphasis on measurement and focused on the processing component of intelligence (Fagan, 2000). The psychometric definition of intelligence has led to debates about what should constitute the focus of IQ tests. Some have argued that the measurement of intelligence is based primarily on the concept of *g* (general intelligence) and related subabilities, whereas other researchers have posed that intelligence should be measured as numerous intelligences of more or less equal status (e.g., Gardner, 1999). The focus on *g* has been predominant in the literature and has proved to be one of the most controversial issues in psychology with respect to race (Deary, 2000).

As Rushton and Jensen (2005) concede, researchers have challenged the derivation of *g* as being a statistical artifact based on factor analysis. Even Spearman’s (1927) early work notes the limitations of *g* as “a hypothetical and purely quantitative factor” (p. 5). Rushton and Jensen cite Spearman’s hypothesis indicating that racial group differences would be largest on *g*-loaded measures. Though they note that particular tasks are more “*g* saturated” than others, the discussion alludes to *g* as a unitary construct in relation to various measures of aptitude. This is clearly not the case, given that standardized IQ tests measure multiple abilities and therefore have differential loadings in relationship to *g*.

In response to accusations that *g* is a statistical “artifact” of factor (or principal-components) analysis, others have noted that “it need not occur. If, in fact, there were mental abilities that were independent of others they would be uncorrelated and they would not load on *g*” (Hayman, 1998, p. 9). While this is true in theory, in practice, new IQ tests that do not correlate with popular measures currently in existence are considered to be problematic in terms of validity. It is clear that among the “best sellers” in the testing domain, the way to validate a new test is by correlating it with other well-established cognitive instruments (Valencia & Suzuki, 2001). Based on this practice, it is unlikely that a measure unrelated to *g* will emerge as a winner in current practice. Thus, it is no wonder that the intelligence hierarchy for different racial/ethnic groups remains consistent across different measures. The tests are highly correlated among each other and are similar in item structure and format. In addition, many predictive validity studies note correlations among IQ, level of education, income, and socioeconomic status. As noted by White (2000), “these are anything but independent variables; they are criteria for one another” (p. 40).

Full scale IQ (FSIQ), g, and racial/ethnic group differences. The FSIQ is the score that is cited as the basis for the racial/ethnic group hierarchy in intelligence, with a mean of 100 and standard deviation of 15 (e.g., Wechsler, 1997). As the literature indicates, the FSIQ is not a pure indicator of *g*. Subtest *g* loadings for different subtests have been found to vary by racial group. Thus, the order of magnitude for *g* loadings for Blacks and Whites can be “considerably unique” (Kaufman, 1990, p. 254). Some may argue that this is unimportant, because regardless of whether a test is a pure measure of *g*, it can still measure something meaningful. Yet given that tests measure more than just *g*, the psychometric definition of intelligence may be challenged and performance on intelligence tests may be more malleable than assumed in past theories.

Genetics and Heritability

In support of their genetic arguments, Rushton and Jensen (2005) cite research documenting results of twin and sibling studies, anatomical differences (e.g., brain size, brain metabolism), processing speed differences, as well as other factors that differentiate between racial groups. However, their either-or method of scoring the evidence between the genetic versus culture-only data implies a misleading dichotomy (Deary, 2000). There are clear interactions among genetic factors, anatomical structures, culture, and environment. The importance of particular interactions may vary depending on an individual’s circumstances and not their racial group membership.

The genetic explanation for the racial/ethnic hierarchy of intelligence is also based largely on estimates of heritability. Heritability estimates are based on correlations of traits between biologically related individuals (Lewontin, Rose, & Kamin, 1984). Most often, correlations are derived from twin and adoption studies. These are limited given that relatives resemble one another because they share genetic traits and live in similar environments. In addition, research on heritability estimates for minority populations is limited, given small sample sizes and geographic regionalism (Suzuki & Valencia, 1997). Thus, the complexities of the culture and genetic interactions make teasing apart the individual contributions of each difficult, if not impossible.

Culture

Over the years, culture has been assigned various definitions. The complexities and ambiguities of the definition of culture are extensive and incorporate multiple levels of meaning across generations (Geertz, 1973).

According to Rushton and Jensen (2005), there are four data sources that are believed to remove the cultural component in support of the genetic argument. These include neurological studies (e.g., reaction time), physiological studies (e.g., anatomical), inheritance studies, and adoption studies. Limitations in these research bases from a cultural–environmental perspective have also been noted in the literature (Hayman, 1998) but are not mentioned by Rushton and Jensen. In particular, the major assumption that differences in culture do not affect these supposedly culture-free measures is questionable. Physiological measures in this case are being used to approximate psychological variables (i.e., intelligence). Evidence supports that culture affects nearly all psychological phenomena; there-

fore, it is entirely possible that biological indicators of intelligence are also affected.

Race

Although Rushton and Jensen (2005) adhere to a biological definition of race, other theorists such as Loury (2001) have emphasized the important social underpinnings of this construct. In this view, although race refers to physical characteristics, the emphasis is placed on the social meanings or interpretations of these features made in society.

If race is, therefore, as much a social category as a biological one, then it would follow that race differences in intellectual performance are not simply mediated by genetics to the exclusion of cultural and environmental factors. The reality is, “under the skin, there is very little order to real human genetic variation” (Cohen, 2002, p. 211). Cohen (2002) noted that of the 15,000 to 20,000 gene pairs that exist, only 6, or 0.03%, are linked to skin color. In addition, it should be noted that skin color and other phenotypic markers are only grossly related to race (Cohen, 2002). Therefore, the associations made by Rushton and Jensen (2005) between race and IQ are questionable.

A related issue with respect to racial group differences in intelligence has been the consistent finding that the variance within racial groups is much greater than that found between racial groups (Valencia & Suzuki, 2001). “Average group differences in *g* are simply aggregated individual differences in *g*, so the composition of racial group differences and individual differences are of the same essential nature” (Jensen, 2000, p. 124). This conclusion, however, has been challenged by Fagan and Holland (2002), whose research suggests that the “average difference of 15 IQ points between Blacks and Whites is not due to the same genetic and environmental factors, in the same ratio, that account for differences among individuals within a racial group in IQ” (p. 382). These results indicate the need to seek further explanations for intelligence differences and to look beyond racially aggregated intelligence test data.

Historical, Contextual, and Testing Issues

Ruston and Jensen (2005) acknowledge in a few sentences the contribution of other theoretical and empirical work supporting an environmental–cultural perspective. These include stereotype threat, mediated learning, and the impact of relative functionalism with respect to particular marginalized groups.

Stereotype Threat

Stereotype threat is defined as anxiety regarding one’s performance in a particular domain (e.g., intelligence) based on negative stereotypes that exist in reference to one’s group (e.g., racial/ethnic group; Aronson, 2002; Steele & Aronson, 1995). This anxiety is not related to the individual’s ability but rather to the situation in which a negative stereotype (e.g., “Blacks are unintelligent”) may be confirmed by one’s performance. Evidence for stereotype threat’s effects is now abundant. Numerous studies show that it can depress the standardized test performances on a variety of groups for whom stereotypes allege inferior abilities in some domain (see Aronson, 2002, for a review).

Rushton and Jensen (2005) minimize the stereotype threat evidence, arguing that it cannot account for cases in which Blacks are in the majority, such as in the sub-Saharan, where despite outnumbering Whites, Blacks perform less well on IQ tests. This work demonstrates that Blacks and Whites experience testing situations differently, often in ways that have a meaningful impact on scores. This effect does not require numerical minority status. Studies have replicated the stereotype threat effect even in all-Black colleges (Aronson, 2002), so it is certainly conceivable that sub-Saharan Blacks could be affected. In addition, Rushton and Jensen ignore the fact that people exist in sociopolitical contexts that have a profound impact on their experience and worldview. Sub-Saharan Blacks operate within a context of racism and colonialism that, in turn, creates and shapes stereotypes. Therefore, when one applies tests constructed by Whites within one cultural context (i.e., American) and then applies them to Blacks and Whites in another, the tests do not mysteriously lose their bias. Stereotype threat may therefore partly explain why any group alleged to be inferior may underperform groups thought to be superior, regardless of their numerical representation in a classroom, in a community, or in a country.

Effects of Mediated Learning Experiences

Studies have also indicated that performance on highly *g*-loaded tasks can be affected through intervention such as exposure to information and dynamic assessment procedures. For example, Skuy et al. (2002) indicated that performance on a highly *g*-loaded task (i.e., Raven's Standard Progressive Matrices [RSPM]) can be improved significantly through mediated learning experiences. Skuy et al. concluded that "African students, by virtue of their sociopolitical history, are especially likely to have been deprived of mediated learning experience" (Skuy et al., 2002, p. 230). Thus, scores on the RSPM may be "more related to schooling, literacy, and the cognitive demands imposed by the environment, and, thus, they may vary more from culture to culture" (Skuy et al., 2002, pp. 230–231). Other studies also indicate that mediated learning interventions were effective in raising the measured indicators of cognitive ability for Black children (see Fagan & Holland, 2002; Sternberg et al., 2002).

Relative Functionalism

Rushton and Jensen (2005) also report findings indicating the relatively high intelligence of Asians in comparison with other racial/ethnic groups. They fail to mention explanations such as relative functionalism that have been used to explain the high achievements of Asians in terms of the educational achievement and the intelligence hierarchy. Relative functionalism suggests that groups will pursue opportunities for achievement in particular contexts (e.g., academic, social, vocational) when it is perceived that other avenues to success are closed. Sue and Okazaki (1990) refuted the notion that Asians are genetically superior to other racial/ethnic groups. On the contrary, they cited relative functionalism as accounting for the high achievement of Asian Americans beyond their measured IQ. This theory posits that Asian Americans experience opportunities for upward mobility in educational areas and exclusion from other noneducational pursuits (e.g., entertainment, politics) because of social discrimination or limited English lan-

guage skills. Though relative functionalism has been difficult to test empirically, anecdotal evidence in terms of the experiences of Asians in the United States seems to support this explanation. Arguments based on relative functionalism could also be made with respect to the limited educational achievements of African Americans due to slavery and historically little access to educational opportunities.

Test Development Practices

Current test development practices have served to maintain the racial/ethnic group hierarchy of intelligence test scores. Strategies used to address issues of cultural bias are limited to expert review panels and various statistical formulations. However, these practices have been criticized on the basis of their conceptual limitations (Valencia & Suzuki, 2001).

Sternberg (2000) criticized current methods of establishing test validity. He noted that intelligence can be represented in terms of a person's talents and the abilities that are valued in a particular sociocultural context. To the extent that one's behavior is discrepant from that valued by society, these individuals will be viewed as less successful and intelligent. Sternberg stated, "tests are validated almost exclusively against the societally approved criteria, giving tests an appearance of validity that they may not have within a given sociocultural group" (Sternberg, 2000, p. 165). Issues of how one adapts to particular environmental contexts that may differ from the status quo are not considered.

With respect to specific tests, accusations of "cultural" and "statistical" bias are still noted for popular tests such as the SAT (Freedle, 2003). Freedle (2003) contended that a corrective scoring method, the Revised-SAT (R-SAT), be used to address the "nonrandom ethnic test bias patterns found in the SAT" (p. 1) by focusing on the "hard" items of the SAT. These hard items are often dependent on "rare vocabulary" (Freedle, 2003, p. 2). Freedle cited work using differential item functioning, which reflects a "small" but "highly patterned nature; that is many *easy* items show a small but persistent effect of African Americans' underperformance, while many *hard* items show their overperformance" (Freedle, 2003, p. 3). Freedle referenced the cultural unfamiliarity hypothesis that "many *easy* verbal items tap into a more culturally specific content and therefore are hypothesized to be perceived differently, depending on one's particular cultural and socioeconomic background" (Freedle, 2003, p. 7). *Hard* items are less ambiguous given that they are most often used in an academic setting. The R-SAT has reduced the Black-White test gap by one third. Verbal scores are particularly affected as Freedle noted that scores on the Verbal R-SAT are increased by as much as 200 to 300 points for individual minority test takers.

Further challenges to the SAT are noted by Rosner (2003), executive director of the Princeton Review Foundation. His research on the 1998 version of the SAT indicates that the percentage of White students answering questions correctly was higher than the percentage of Black students for all 138 items. Items with higher percentages of Black students answering correctly in comparison with Whites were "systematically" rejected during the pretesting phase of the instrument development (Rosner, 2003).

Implications

It is evident that to reach Rushton and Jensen's (2005) position on the meaning of the race differences in test performance, one has to accept a particular definition of intelligence and believe in the validity of IQ tests to measure it. There is also growing documentation of the powerful effects of context on intellectual performance (e.g., stereotype threat) and learning (e.g., mediated learning). Even admitting the possibility of racially based differences in intelligence, there appears to be considerable research supporting environmental/cultural justification for race differences—enough at least to make one question a steadfast belief in a biological explanation.

It appears that the culture versus genetic debate will continue despite the fact that most would adhere to an interactionist perspective (Reynolds, 2000). As noted in the beginning of this article, our concerns focus on the implications of the genetic argument. Where society stands on the malleability of intelligence will affect the allocation of resources (e.g., affirmation action) and the promotion of particular methods of intervention (e.g., educational programs like Head Start).

Our commentary has only briefly highlighted the literature with respect to possible cultural and environmental explanations for the racial/ethnic group hierarchy on intelligence tests. The theoretical and empirical work appear promising in this area. In addition, questions may be raised regarding current test development practices (from item selection to validation). There appears to be many opportunities to think “outside the box” in our examination of what constitutes an intelligence measure and how we examine issues of bias (White, 2000). In addition, given growing concerns regarding the usage of intelligence tests for selection purposes, Jensen (2000) suggested using criteria that go beyond standardized measures and the inclusion of indicators of past performance (e.g., work history). The goal for all of us is to discover “truth” in whatever form it may take. Reynolds (2000) called for members of the profession to base interpretations of racial differences on mental tests on empirical data and continually challenge assumptions about the meaning of these differences.

It appears that many challenges remain in explaining fully the racial/ethnic group hierarchy of intelligence whether one adheres to the culture-only or genetic perspective. We believe that the answer resides most likely in the interaction between the two and that data supporting the malleability of IQ will prevail.

References

- Aronson, J. M. (Ed.). (2002). *Improving academic achievement: Impact of psychological factors in education*. San Diego, CA: Academic Press.
- Cohen, M. N. (2002). An anthropologist looks at “race” and IQ testing. In J. M. Fish (Ed.), *Race and intelligence: Separating science from myth* (pp. 201–224). Mahwah, NJ: Erlbaum.
- Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain*. New York: Oxford University Press.
- Fagan, J. F. (2000). A theory of intelligence as processing: Implications for society. *Psychology, Public Policy, and Law*, 6, 168–179.
- Fagan, J. F., & Holland, C. R. (2002). Equal opportunity and racial differences in IQ. *Intelligence*, 30, 361–387.

- Freedle, R. O. (2003). Correcting the SAT's ethnic and social-class bias: A method for reestimating SAT scores. *Harvard Educational Review, 73*, 1–42.
- Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century*. New York: Basic Books.
- Geertz, C. (1973). *The interpretation of cultures*. New York: Basic Books.
- Hayman, R. L., Jr. (1998). *The smart culture: Society, intelligence and law*. New York: New York University Press.
- Jensen, A. R. (2000). Testing: The dilemma of group differences. *Psychology, Public Policy, and Law, 6*, 121–127.
- Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence*. Boston: Allyn & Bacon.
- Lewontin, R. C., Rose, S., & Kamin, L. (1984). *Biology, ideology, and human nature: Not in our genes*. New York: Pantheon Books.
- Loury, G. C. (2001). *The anatomy of racial inequality*. Cambridge, MA: Harvard University Press.
- Reynolds, C. R. (2000). Why is psychometric research on bias in mental testing so often ignored. *Psychology, Public Policy, and Law, 6*, 144–150.
- Rosner, J. (2003, April 14). On White preferences. *Nation, 276*, 24.
- 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.
- 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.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. *Journal of Personality and Social Psychology, 69*, 797–811.
- Sternberg, R. J. (1996). Myths, countermyths, and truths about intelligence. *Educational Researcher, 25*, 11–16.
- Sternberg, R. J. (2000). Implicit theories of intelligence as exemplar stories of success: Why intelligence test validity is in the eye of the beholder. *Psychology, Public Policy, and Law, 6*, 159–167.
- Sternberg, R. J., Grigorenko, E. L., Ngorosho, D., Tantufuye, E., Mbise, A., Nokes, C., et al. (2002). Assessing intellectual potential in rural Tanzanian school children. *Intelligence, 30*, 141–162.
- Sue, S., & Okazaki, S. (1990). Asian American educational achievements: A phenomenon in search of an explanation. *American Psychologist, 45*, 913–920.
- Suzuki, L., & Valencia, R. R. (1997). Race-ethnicity and measured intelligence: Educational implications. *American Psychologist, 52*, 1103–1114.
- Valencia, R. R., & Suzuki, L. A. (2001). *Intelligence testing and minority students: Foundations, performance factors and assessment issues*. Thousand Oaks, CA: Sage.
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third edition*. San Antonio, TX: Psychological Corporation.
- White, S. H. (2000). Conceptual foundations of IQ testing. *Psychology, Public Policy, and Law, 6*, 33–43.