

## HISTORY OF CULTURAL BIAS IN ASSESSMENT

Philip Sayegh, PhD, MPH

Understanding the history of cultural bias in psychological assessment is crucial given the historical connection between psychological assessment (e.g., intelligence testing) and racist and eugenicist ideology. Indeed, it is well known that the early history of psychological and intelligence testing in the United States was strongly rooted in and remains tarnished by eugenics, blatant ethnic and racial discrimination, and attempts to oppress individuals who were not from the dominant culture (e.g., Allen, 2006; Franklin, 2007; Guthrie, 2004; Sayegh et al., 2023). Contemporary Euro-American psychological assessment is, unfortunately, shaped by inherent biases of the dominant Western culture, such as the narrow racial, cultural, and social identities and norms informing it. In addition, the early functions of psychological assessment in Europe and the United States (U.S.) in the 19th century reflected sociocultural interests of the time (e.g., racial segregation in schools [Terman, 1916] and immigration restriction [Allen, 2006]). The consequent 20th-century testing movement similarly continued to rely on definitions of intellectual abilities reflective of the dominant culture and measurements (e.g., norms) based on populations from the dominant culture. Thus, inherent racial and cultural bias, as well as social injustice, were inevitable (e.g., Sayegh et al., 2023).

## EARLIER EFFORTS TO ASSESS INTELLECTUAL FUNCTIONING

It is believed that the initial prevalent use of psychological assessment took place over 3,000 years ago in ancient China, which often involved tests of abilities such as visuospatial perception, problem solving, and divergent thinking, as well as creativity. During China's Sui dynasty (581–618 B.C.), a Civil Service Examination system was created to help the emperor identify employees and evaluate his bureaucrats' suitability to their positions (Zhang & Fang, 2020). This examination measured cultural knowledge, planning/administration, and martial arts (Wang, 1993). These types of examination were modified but continued to be used in China until the early part of the 21st century, though very few, if any, other nations sought to replicate these or create other methods for assessment on a national level until the 20th century (Oakland et al., 2001). What is notable about these earliest assessments is the strong emphasis on culture (e.g., cultural and Confucian classics knowledge; Zhang & Fang, 2020).

As described by Allen (2006), in the late 18th to late 19th century, the assessment of mental abilities was conducted by employing craniometry, which was thought to be a physiological indicator of intellectual capacity, and these biased criteria were applied to justify the mental superiority of non-Latine White men above women and non-White groups. Moreover, Allen reported that findings from craniometries were used

### REMEMBER

From its earliest days, psychological assessment was culturally biased. In the United States and Europe, early efforts to measure intelligence emphasized physical attributes and were biased toward non-Latine White men, and results were often used to oppress minoritized individuals.

in Europe in an effort to thwart women's suffrage and university admissions. Meanwhile, in both the United States and Europe, such findings were used to oppose abolishing slavery. Sir Francis Galton (1822–1911), who coined the term “eugenics,” created the first broad “intelligence” test in the late 1800s focused on the measurement of other physical attributes, such as visual acuity, reaction time, and grip strength (Hunt, 2009).

## ENTERING THE MODERN ERA OF INTELLIGENCE TESTING

In a shift from craniometry and the measurement of physical attributes, contemporary intelligence tests were beginning to be developed in the early 20th century. The first modern intelligence test, the Binet-Simon intelligence scale, was developed in 1905 by Alfred Binet (1857–1911) and his collaborator, Théodore Simon (1873–1961), and they later revised this test in 1908 and 1911. The purpose of the test, as commissioned by the French government, was to assist in the identification of students who may have benefited from remedial studies or special

education. Binet viewed the test as a measure of what a child has learned up to their particular age rather than intelligence or one's innate capacity to learn (Allen, 2006). Moving forward, however, most intelligence tests in the United States were originally used or created in the early 20th century solely to establish the supremacy of non-Latine Whites compared to individuals from other racial and ethnic groups, immigrants from certain countries, and other minoritized groups, as detailed by Guthrie (2004) in *Even the Rat Was White: A Historical View of Psychology* (2nd ed.) and Franklin (2007).

## **Intelligence Testing as a Means to Oppress Minoritized Groups**

### ***Youth and Academic Settings***

Lewis Terman (1877–1956), the noted eugenicist who published the Stanford-Binet Intelligence Scales in 1916, asserted that individuals from racial and ethnic minority groups (i.e., Native Americans, Mexican Americans, and African Americans) were intellectually inferior, which he attributed to race and genetic factors, and he called for segregation of students accordingly. He based this assertion in part on results from the Stanford-Binet Test administered to students in the United States that showed that Black students scored lower than non-Latine White students (and female students scored lower than male students; Terman, 1916). These results suggesting the superiority of non-Latine Whites in terms of intelligence were widely publicized (Guthrie, 1998). Troublingly, in the 1936 revision of this test, the test was adjusted to account for the male-female difference but not the racial-group difference (Allen, 2006).

Terman's (1916) assertion and findings regarding racial differences in and hereditary influences on intelligence were discordant with findings from an earlier study conducted by Stetson (1897) that demonstrated that Black students (mean age of 11 years) slightly outperformed their non-Latine White counterparts in memorization capabilities. Despite the findings showing stronger mnemonic performance among the Black students in the study, Stetson linked these strengths to "primitive brains" and further stated that Black youth were "deficient in reasoning" (American Psychological Association [APA], 2023, p. 4). Moreover, given that this study's results did not provide support for the predominant conviction in the superiority of non-Latine White individuals, critics suggested that the test be revised, and they discounted the findings (Guthrie, 1998).

Similarly, around this same time in England, Sir Cyril Burt (1883–1971) conducted research that led him to conclude that intelligence was primarily hereditary and that those (e.g., schoolchildren) of European background possessed superior intelligence compared to others (Hearnshaw, 1979). However, it is now widely known that Burt likely fabricated his data to support this eugenicist view (e.g., Scarr, 1994).

### ***Xenophobia and Eugenics: Immigration, Deportation, and Sterilization***

Intellectual and cognitive testing were also employed in an effort to prevent immigration to the United States through Ellis Island, New York (e.g., by Henry H. Goddard [1866–1957], an American eugenicist who held strong beliefs pertaining to the hereditary nature of intelligence) during the World War I era. The results from such tests were both inaccurately and inappropriately used to make false generalizations about whole immigrant groups, such as Southern Europeans and those of Jewish origin (with the vast majority of individuals from these groups being classified as mentally deficient based on the testing), which caused a sharp increase (400% increase between 1913 and 1914) in deportations and influenced the passing of immigration restriction legislation (Allen, 2006).

Similarly troubling, in 1914, Goddard served on the Committee to Study and Report on the Best Practical Means of Cutting Off the Defective Germ-Plasm in the American Population as the representative for psychology. Ultimately, the Committee advocated for both segregation and sterilization of those deemed mentally inferior and/or socially inept. The Committee summoned the field of psychology to help establish tests and standards to assist in the identification of these individuals (Schuster, 1914), which resulted in the sterilization of 35,000 individuals in the United States (Greenwood, 2017). A large proportion of these people were immigrants, Black, indigenous, and poor non-Latine White, and/or had disabilities (Kevles, 1968).

### ***Race, Geography, and Military Settings***

Just a few years later, in 1917, then-APA president Robert M. Yerkes (1876–1956) introduced the Committee on the Psychological Examination of Recruits, which called for every U.S. soldier to be mentally examined (APA, 2023). Results indicated that individuals of northern European, Anglo-Saxon, and Nordic descent scored significantly higher than their Southern and Eastern European and Mediterranean counterparts (Allen, 2006). The Committee commissioned by Yerkes also found racial-group differences on their tests. However, as noted in an APA *Historical Chronology* focused on racial and cultural hierarchy and inequity in the United States (APA, 2023), “Culturally

biased test questions and examination procedures, along with an assumption of White, American normativity, led to results that established a clear racial hierarchy of performance with White, American-born recruits scoring highest on IQ tests and Black recruits, particularly those from the American south, scoring the lowest” (p. 6). Consequently, the vast majority (89%) of Black recruits were classified as “morons” by the *Committee* (e.g., Rury, 1988).

## **REMEMBER**

Even in more contemporary times, biased intelligence testing has been used to oppress individuals from minoritized groups. The oppression has been wide ranging and includes school segregation and sterilization by race, as well as unjust immigration restriction and deportation.

## Contemporary Examples of Cultural Bias and Oppression in Assessment

### *Intelligence Testing, Public Schools, and the Law*

In 1954, the *Brown v. Board of Education* Supreme Court case ruling outlawed segregation of public schools by race. Shortly after this landmark ruling, Audrey Shuey (1900–1977) used early intelligence testing results in an effort to counter this ruling and call for the segregation of students in U.S. schools by race in the book she published titled *The Testing of Negro Intelligence* (1958). Shuey reviewed and provided a summary of results of intelligence testing scores among Black and non-Latine White individuals between the 1910s and 1960s across the United States as well as in Jamaica and Canada that indicated the persistence of a 15-point difference in IQ scores across groups in favor of non-Latine White individuals, which she attributed to “native differences between Negroes and whites” (p. 318). While these test score differences are indeed present, such hereditary interpretations fail to address the strong influence of environmental (such as oppression, educational access, and inequity) and other factors on intellectual functioning (e.g., Turkheimer et al., 2003; Weiss & Saklofke, 2020).

Beginning in the late 1960s, a series of judicial decisions was made pertaining directly to cultural and racial bias in assessment. *Hobson v. Hansen* (1967) represents the initial significant case that called special education placement practices into question. The ruling for this case stated that the grouping of students into tracks based on test scores was in violation of constitutional rights, as it discriminated against Black and poorer individuals. During the Association of Black Psychologists national meeting in 1968, a call was issued for a moratorium on IQ and other standardized testing for Black students given the bias and its detrimental consequences (Auguste & Griffin, 2020).

In 1969, *Diana v. State Board of Education* (1970) was filed in a Northern District of California court. This case also pertained to concerns over the use of standardized tests for students’ academic placements. Specifically, nine Spanish-speaking Latine students were administered English-language IQ tests (i.e., Stanford-Binet Intelligence Test or the Wechsler Intelligence Scale for Children) and were classified as intellectually disabled, resulting in special educational placement. In 1970, the case was settled out of court by consent decree, and it resulted in state law requirements for native language testing of students to help avoid special education placement solely due to limited English-language proficiency. Not long after, the *Larry P. v. Riles* case, filed in 1971 in Northern California and settled in 1979, expanded on the *Diana v. State Board of Education* (1970) ruling by requiring schools to administer tests that do not discriminate based on racial background to help prevent inappropriate special education placement.

However, in a step backward, the *Parents in Action on Special Education (PASE) v. Hannon* (1980) judge ruled that the level of racial bias in IQ tests was insignificant and that special education placements should be maintained for the Black students in

Illinois who inspired the case. To date, California remains the only state that has a ban on IQ testing for Black students in public schools for the purpose of special education placement, which was fostered by the *Larry P. v. Riles* decision followed by additional court rulings and intervention by the California Department of Education. California is also currently the only U.S. state that prohibits school psychologists from assessing Black students using IQ tests (Frisby & Henry, 2016). Although seemingly well-intentioned in its aim to correct the disproportionate placement of Black youth in special education classrooms, it has in fact become a barrier to eligibility for accessing needed special education services that students are legally entitled to under IDEA, thereby further contributing to disenfranchisement. As Dalton (2023) wrote, “While an IQ test alone should never be determinative, the lack of one is regularly cited as a reason why a student who is black cannot qualify under some disability categories. Specific Learning Disability is but one such example. This suggests that even the best of intentions can go awry. Today, black students in California public schools do not have the same opportunity for assessment that other races enjoy” (p. 2).

### ***Popular Literature***

More recently, in the 1994 book titled *The Bell Curve: Intelligence and Class Structure in American Life* by psychologist Richard J. Herrnstein (1930–1994) and political scientist Charles Murray (1943–; Herrnstein & Murray, 1994), the authors stated that intelligence tests are not biased against minoritized ethnoracial groups and added that group differences in intelligence quotient (IQ) scores are valid. Moreover, the authors stated that intelligence is impacted primarily by hereditary influences. While it is now known that hereditary factors do account for approximately 50% of the variance in individual levels of estimated intellectual abilities (Plomin & von Stumm, 2018), it is unwarranted and harmful that these authors and others cite this as evidence to advocate for policy changes such as the elimination of welfare policies while discounting the key role of environmental factors and social influences, such as socioeconomic status, educational opportunity, life stressors, and early adversity, on neurocognitive development (e.g., Turkheimer et al., 2003; Weiss & Saklofke, 2020).

### ***Medical-Legal Assessments***

As an example of the ongoing presence of cultural bias in assessment, there has been a recent troubling development pertaining to the neurocognitive testing of former National Football League (NFL) players with a history of concussion(s). These cases resulted in a \$765 million legal settlement in 2013, described in a New York Times article by Ken Belson titled “Black Former NFL Players Say Racial Bias Skews Concussion Payouts” (2020, August 25). Specifically, retired NFL players are alleging explicit and deliberate discrimination with regard to their dementia-related claims due to the use of African American-specific norms that result in the need to demonstrate “steeper cognitive declines to qualify for a payout.” Although at face value it may appear appropriate or perhaps even well-intentioned to utilize normative data specific to African Americans for cognitive test results, this practice is not without

limitations and has been deemed questionable in the literature (e.g., Fields et al., 2013; Manly, 2005; Rohling et al., 2015). For example, Manly and Echemendia (2007) reported that while many have suggested that the use of race-specific norms can improve our tests' sensitivity and specificity in identifying cognitive impairment, this practice carries its own notable limitations, namely that the norms do not account for the "underlying cultural and educational factors for which race serves as a proxy" and that "setting 'more lenient' cutoffs for impairment among ethnic minorities denies these groups needed services." Moreover, it is important to note that these normative data are derived from the general population, not unique groups like NFL players, nearly all of whom have been enrolled in college. They also do not take into account individuals who identify as biracial/multiracial or the diversity among individuals who identify as Black.

In practice, this means Black players have to show steeper cognitive declines to qualify for a payout. This raises concerns that race is being used to determine whether individuals will receive benefit. Indeed, the Belson (2020, August 25) article includes a quote from Thomas Berg, J.D., a professor of law and public policy, stating, "If a racial factor is being used against a historically disadvantaged group to deny benefits that they would otherwise receive, that is illegal." Additionally, Robert Stern, PhD, a neuropsychologist and professor of neurology, is quoted in the same article, stating, "...when [these norms] are used across the board in an algorithm to determine monetary compensation, it is inappropriate and results in injustice and racial inequities." Needless to say, this recently discovered finding is disturbing on multiple levels, as the practices used seemingly "revive" many of the earliest key problems with testing and systemic discrimination that the field should be striving to eradicate.

### ***Sexual and Gender Minoritized Persons***

As is the case for racial, cultural, and ethnic minoritized populations, the history of psychological assessment among sexual minoritized (gay men, lesbians, bisexuals, and other non-heterosexual individuals) and gender minoritized (transgender, gender nonbinary, gender expansive, and other diverse gender representations) persons (LGBTQIA+) is marred by bias and oppression. While bias against other various minoritized groups had been established earlier (e.g., Walsh & Betz, 1995), Chernin and colleagues (1997) noted that it was not until the publication of their work specifically focused on heterosexism in assessment that bias in assessment among LGBTQIA+ individuals had been addressed in the literature, based on their review of the psychological literature from 1974 to the time of their writing. In a published response to this article, Prince (1997) highlighted that although some book chapters had in fact been previously published focused on psychological testing and sexual orientation (e.g., Gonsiorek, 1982; Pope, 1992; Prince, 1995), "The virtual absence of this topic throughout the professional literature is testimony to the pervasiveness of such bias, and to the insidious power of that bias to operate through omission ... Unfortunately, as Chernin et al. clearly pointed out, specific lack of

attention to sexual orientation has allowed continued use of a variety of instruments whose scales and manuals include offensive, exclusionary, and stigmatizing language” (p. 82).

Chernin and colleagues (1997) noted that bias in psychological assessment for LGBTQIA+ persons can be present for a whole host of reasons, such as the presence of and wording of items in questionnaires and inventories that may not be applicable (thus precluding meaningful interpretation), omission bias (e.g., lack of references to same-sex parents in the Minnesota Multiphasic Personality Inventory [MMPI] items), direct or indirect references to psychopathology among this group, including connotation bias (e.g., Scale 5 [Masculinity-Femininity] on the MMPI-2), scales meant to diagnose mental disorders amongst scales meant to characterize minoritized groups (contiguity bias; e.g., reference to homosexuality in the MMPI-2 Scale 5 interpretation), and stereotypes among providers that may confound interpretation (even when bias is limited).

Moreover, there is a deficiency of measures normed and validated specifically on these populations, and assessment items and scales have not been created to appropriately assess the full span of experiences and concerns that contextualize the lives of LGBTQIA+ individuals (Prince, 1997). As Gonsiorek (1982) noted, results from earlier studies involving psychological assessment of gay men and lesbians have been severely flawed due to sampling and other methodological issues, misinterpretation of differences in test scores across groups falling within the normal range, and a lack of appreciation of base rates (e.g., related to differences in environmental stress). These biases can render the tools themselves harmful, as they may lead to damaging, erroneous conclusions that further oppress LGBTQIA+ people by, for example, reinforcing negative stereotypes about and pathologizing individuals from this group (Gonsiorek 1982; Prince, 1997).

Given the longstanding history of pathologizing and stigmatizing transgender and gender expansive persons both within the field of psychology as well as society at large (Glicksman, 2013), often to a greater extent than sexual minoritized persons (Norton & Herek, 2013), bias in assessment is particularly problematic and harmful for this group. For instance, psychological measures that are normed by sex or gender lack norms for transgender persons and are not validated among this group (Keo-Meier & Fitzgerald, 2017; Lewitzke, 2018), which can result in erroneous interpretations and conclusions that suggest psychopathology rather than “. . . a manifestation of physical, social, and psychological conflict with which the client is enduring” (Lewitzke, 2018, p. 81). Similar to sexual minoritized individuals, years of research involving various iterations of the MMPI and other psychological tests among transgender persons have suggested a high prevalence of psychopathology and psychiatric comorbidity (e.g., Hepp et al., 2005). However, these (likely biased) findings were not interpreted through an appropriate and current theoretical lens such as gender minority stress (Hendricks & Testa, 2012). Further, it is wholly inadequate to state that transgender and gender expansive individuals can adequately and accurately be assessed by a test

that uses non-gendered norms, as these norms are almost (if not entirely) exclusively comprised of cisgender men and women. Transgender and gender diverse individuals are not represented in these normative samples, and thus conclusions drawn from scores are often erroneous.

As a result of such bias in psychological assessment among transgender and gender expansive persons, often resulting in misinformed and harmful interpretations, these clients can face numerous negative consequences, such as loss of child custody and job opportunities, as well as restriction or preclusion of access to gender-affirming care (Keo-Meier & Fitzgerald, 2017). Regarding gender-affirming care, starting in the late 1970s, standards of care were beginning to be developed that resulted in a gatekeeping model (Lev, 2005) wherein transgender persons were required to prove that they were in fact transgender and pass tests established by clinicians (Keo-Meier & Fitzgerald, 2017). Despite some progress in the field in terms of research focused on assessing gender dysphoria in youth (e.g., Leibowitz & Telingator, 2012), "... there is no consensus among clinicians today regarding the role of psychological tests in evaluations for hormone and surgical treatment of gender transition in adults... The practice of psychological testing using tests that are not normed on this population or for these purposes must be scrutinized" (Keo-Meier & Fitzgerald, 2017, p. 52). There is an ongoing need to strive to reduce bias in assessment to help improve access to medical and surgical interventions that have been shown to be associated with better mental health outcomes in transgender persons (e.g., Baker et al., 2021; Gómez-Gil et al., 2012; Keo-Meier et al., 2015).

## REMEMBER

It is inadequate to say a test has non-gendered norms and so transgender and gender expansive individuals can accurately be assessed by it. Non-gendered norms typically mean a normative sample comprised almost (if not entirely) exclusively of cisgender men and women. Transgender and gender diverse individuals are not represented in these normative samples, and thus conclusions drawn from scores are often erroneous.

### *Research*

Researchers as well as peer reviewers and editors of scientific journals should be mindful of the ongoing bias in studies that measure intelligence and neurocognition in individuals from culturally minoritized background, as this can lead to the publication and dissemination of misleading research findings that promote harmful stereotypes and reinforce racial and social injustice. For example, biased intelligence test findings have led to erroneous, sweeping conclusions regarding the association between population-level IQ and national economic success. More specifically, Lynn and Vanhanen (2002) proposed that differences in economic success across countries are primarily accounted for by their populations' IQ, followed by social influences (i.e., poorer countries, such as many in Africa, are less wealthy than other countries

due to lower IQs in the population). Their correlation findings did not fully support this contention, and these authors did not consider several potentially confounding factors above and beyond test bias, such as historical influences including imperialism and colonization; as Wiggan (2007) duly noted, “Ultimately, they adopted an absolute and universal notion of intelligence created by a few and then applied to the entire world” (p. 314).

A recent troubling example of this problem is seen in a now-retracted study conducted by Clark and colleagues (2020) that investigated the association between socioeconomic indicators and estimated average intellectual ability (i.e., IQ) in 140 countries. The data reported by these authors indicated that many African, Central American, and South Asian countries have a mean IQ below 50, and that adults from African countries had an average IQ estimate that is approximately 1.6 standard deviations lower than that of adults from European countries. In response to these data, Ebbesen (2020) issued a written critique of these findings, stating, “These notions are incompatible with psychological science and all conclusions drawn from these data are invalid” (p. 1). More specifically, Ebbesen stated that, “. . .the methods used to estimate national IQ are flawed and unreliable” due to various methodological issues such as “the use of nonrepresentative samples, biased inclusion and exclusion of data, and miscalculations” (p. 2). Ebbesen recommended that Clark and colleagues’ paper be retracted. He also provided a practical suggestion to help prevent the publication of inappropriate psychological test findings that can reinforce systemic racial and cultural injustice, recommending that journal editors “. . .mandate that authors not just provide summary statistics in tables, but also include plots showing the underlying raw data points” (p. 3).

Recently, Buchanan and colleagues (2021) provided a wide-ranging collection of recommendations aimed at reducing racism in the conduct, reporting, reviewing, and dissemination of psychological science. Their suggestions include, but are not limited to, the utilization of the most appropriate research methods for minoritized populations, reporting of ethnicity for all research participants as well as the heterogeneity within minoritized populations, increasing the quantity of journal reviewers and editors who identify as coming from a minoritized group, and collaborating with community partners of minoritized populations for the dissemination of relevant research.

## CONCLUSION

The history of cultural bias in assessment is not quite history yet. Its detrimental impacts continue to pervade the field in multiple contexts (e.g., academic, clinical, medical-legal, and research) and can lead to ongoing oppression of individuals from minoritized groups. Despite some progress within the field of psychology and judicial rulings aimed at reducing the bias and its detrimental consequences, there remains significant room for improvement, which is urgently needed (e.g., Sayegh et al., 2023).

The information provided in the remainder of this book can serve as an ideal resource for clinicians focused on how to address and account for cultural bias in assessment, both broadly and within specific cultural groups.

## REFERENCES

- Allen, G. E. (2006, September). Intelligence tests and immigration to the United States, 1900–1940. *Encyclopedia of life sciences*. Wiley. <https://doi.org/10.1002/9780470015902.e0005612>
- American Psychological Association (2023). *Historical chronology. Examining psychology's contributions to the belief in racial hierarchy and perpetuation of inequality for people of color in U.S.* Retrieved June 1, 2023 from <https://www.apa.org/about/apa/addressing-racism/historical-chronology.pdf>
- Auguste, Evan E., & Griffin, B. M. (2020). Towards an eternity: Celebrating the Association of Black Psychologists' 50th anniversary. *Psychology from the Margins*, 2, Article 2. Retrieved June 16, 2023, from <https://ideaexchange.uakron.edu/psychologyfromthemargins/vol2/iss1/2>
- Baker, K. E., Wilson, L. M., Sharma, R., Dukhanin, V., McArthur, K., & Robinson, K. A. (2021). Hormone therapy, mental health, and quality of life among transgender people: A systematic review. *Journal of the Endocrine Society*, 5(4), 1–16. <https://doi.org/10.1210/jeandso/bvab011>
- Belson, K. (2020, August 25). Black former NFL players say racial bias skews concussion payouts. *The New York Times*. <https://www.nytimes.com/2020/08/25/sports/football/nfl-concussion-racial-bias.html#:~:text=Two%20retired%20players%20have%20accused,as%20much%20as%20%24%20million>
- Brown v. Board of Education*, 347 U.S. 483 (1954).
- Buchanan, N. T., Perez, M., Prinstein, M. J., & Thurston, I. B. (2021). Upending racism in psychological science: Strategies to change how science is conducted, reported, reviewed, and disseminated. *American Psychologist*, 76(7), 1097–1112. <https://doi.org/10.1037/amp0000905>
- Chernin, J., Holden, J. M. & Chandler, C. (1997). Bias in psychological assessment: Heterosexism. *Measurement and Evaluation in Counseling and Development*, 30(2), 68–76. <https://doi.org/10.1080/07481756.1997.12068922>
- Clark, C. J., Winegard, B. M., Beardslee, J., Baumeister, R. F., & Shariff, A. F. (2020). RETRACTED: Declines in religiosity predict increases in violent crime—but not among countries with relatively high average IQ. *Psychological Science*, 31(2), 170–183. <https://doi.org/10.1177/0956797619897915>
- Dalton, M. A. (2023, August 10). *Historical disproportional placement of students in special education based on race and ethnicity*. San Diego Legal Studies Paper No. 23-026. <https://ssrn.com/abstract=4537611> or <http://dx.doi.org/10.2139/ssrn.4537611>
- Diana v. California Board of Education*, No. C-70-37. N.D. Calif. (1970).
- Ebbesen, C. L. (2020). Flawed estimates of cognitive ability in Clark et al. psychological science, 2020. *PsyArXiv*. <https://doi.org/10.31234/osf.io/tzr8c>
- Fields, K., Hill, B., Corley, E., Russ, K., Boettcher, A., Musso, M., & Rohling, M. (2013). 1, 2, 3, 4 . . . I declare norms war! Comparison of the Mitrushina meta-norms, Heaton, and

- comprehensive neuropsychological normative system demographic norms in a matched African American and White sample. *Archives of Clinical Neuropsychology*, 28(6), Article 522.
- Franklin, V. P. (2007). The tests are written for the dogs: African American children, and the intelligence testing movement in historical perspective. *The Journal of Negro Education*, 76(3), 216–229.
- Frisby, C. L., & Henry, B. (2016). Science, politics, and best practice: 35 years after Larry P. *Contemporary School Psychology*, 20(1), 46–62. <https://doi.org/10.1007/s40688-015-0069-3>
- Glicksman, E. (2013). Transgender today. *Monitor on Psychology*, 44. Retrieved August 23, 2023 from <https://www.apa.org/monitor/2013/04/transgender>
- Gómez-Gil, E., Zubiaurre-Elorza, L., Esteva, I., Guillamon, A., Godás, T., Almaraz, M. C., ... & Salamero, M. (2012). Hormone-treated transsexuals report less social distress, anxiety and depression. *Psychoneuroendocrinology*, 37(5), 662–670. <https://doi.org/10.1016/j.psyneuen.2011.08.010>
- Gonsiorek, J. C. (1982). Results of psychological testing in homosexual populations. In W. Paul, J. Weinrich, J. Gonsiorek, & M. Hotvedt (Eds.), *Homosexuality: Social, psychological, and biological issues* (pp. 57–70). Sage.
- Greenwood, J. (2017). Psychologists go to war. *Behavioral Scientist*. <https://behavioralscientist.org/psychologists-go-war/>
- Guthrie, R. V. (1998). *Even the rat was White: A historical view of psychology*. Allyn and Bacon.
- Guthrie, R. V. (2004). *Even the rat was white: A historical view of psychology* (2nd ed.). Pearson Education.
- Hearnshaw, L. S. (1979). *Cyril Burt: Psychologist*. Cornell University Press.
- Hendricks, M. L., & Testa, R. J. (2012). A conceptual framework for clinical work with transgender and gender nonconforming clients: An adaptation of the Minority Stress Model. *Professional Psychology: Research and Practice*, 43(5), 460–467. <https://doi.org/10.1037/a0029597>
- Hepp, U., Kraemer, B., Schnyder, U., Miller, N., & Delsignore, A. (2005). Psychiatric comorbidity in gender identity disorder. *Journal of Psychosomatic Research*, 58(3), 259–261. <https://doi.org/10.1016/j.jpsychores.2004.08.010>
- Herrnstein, R. J., & Murray, C. (2010). *The bell curve: Intelligence and class structure in American life*. Simon and Schuster.
- Hobson v. Hansen*, 269 F. Supp. 401 D.C.C. (1967).
- Hunt, M. (2009). *The story of psychology*. Random House.
- Keo-Meier, C. L., & Fitzgerald, K. M. (2017). Affirmative psychological testing and neurocognitive assessment with transgender adults. *Psychiatric Clinics*, 40(1), 51–64. <http://dx.doi.org/10.1016/j.psc.2016.10.011>
- Keo-Meier, C. L., Herman, L. I., Reisner, S. L., Pardo, S. T., Sharp, C., & Babcock, J. C. (2015). Testosterone treatment and MMPI–2 improvement in transgender men: A prospective controlled study. *Journal of Consulting and Clinical Psychology*, 83(1), 143–156. <https://doi.org/10.1037/a0037599>
- Kevles, D. J. (1968). Testing the army's intelligence: Psychologists and the military in World War I. *The Journal of American History*, 55(3), 565–581. <https://doi.org/10.2307/1891014>
- Larry P. v. Riles*, 495 F. Supp. 926, 987 N.D. Cal. (1979).
- Leibowitz, S. F., & Telingator, C. (2012). Assessing gender identity concerns in children and adolescents: Evaluation, treatments, and outcomes. *Current Psychiatry Reports*, 14, 111–120. <https://doi.org/10.1007/s11920-012-0259-x>

- Lev, A. I. (2005). Disordering gender identity: Gender identity disorder in the DSM-IV-TR. *Journal of Psychology & Human Sexuality, 17*(3–4), 35–69. [https://doi.org/10.1300/J056v17n03\\_03](https://doi.org/10.1300/J056v17n03_03)
- Lewitzke, K. (2018). Working with transgender clients: Considerations for psychological testing and assessment. *Graduate Student Journal of Psychology, 17*, 78–87. <https://doi.org/10.52214/gsjp.v17i.10922>
- Lynn, R., & Vanhanen, T. (2002). *IQ and the wealth of nations*. Praeger p.
- Manly, J. J. (2005). Advantages and disadvantages of separate norms for African Americans. *The Clinical Neuropsychologist, 19*(2), 270–275. <https://doi.org/10.1080/13854040590945346>
- Manly, J. J., & Echemendia, R. J. (2007). Race-specific norms: Using the model of hypertension to understand issues of race, culture, and education in neuropsychology. *Archives of Clinical Neuropsychology, 22*(3), 319–325. <https://doi.org/10.1016/j.acn.2007.01.006>
- Norton, A. T., & Herek, G. M. (2013). Heterosexuals' attitudes toward transgender people: Findings from a national probability sample of U.S. adults. *Sex Roles: A Journal of Research, 68*(11–12), 738–753. <https://doi.org/10.1007/s11199-011-0110-6>
- Oakland, T., Poortinga, Y. H., Schlegel, J., & Hambleton, R. K. (2001). International Test Commission: Its history, current status, and future directions. *International Journal of Testing, 1*(1), 3–32. [https://doi.org/10.1207/S15327574IJT0101\\_2](https://doi.org/10.1207/S15327574IJT0101_2)
- Parents in Action on Special Education (PASE) v. Hannon*, 506 F. Supp. 831 N.D. Ill. (1980).
- Plomin, R., & Von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics, 19*(3), 148–159. <https://doi.org/10.1038/nrg.2017.104>
- Pope, M. (1992). Bias in the interpretation of psychological tests. In S. Dworkin & F. Gutierrez (Eds.), *Counseling gay men and lesbians: Journey to the end of the rainbow* (pp. 277–292). American Counseling Association.
- Prince, J. P. (1995). *Strong Interest Inventory resource: Strategies for group and individual interpretations in college settings*. Consulting Psychologists Press.
- Prince, J. P. (1997). Assessment bias affecting lesbian, gay male and bisexual individuals. *Measurement and Evaluation in Counseling and Development, 30*(2), 82–87. <https://doi.org/10.1080/07481756.1997.12068924>
- Rohling, M. L., Miller, R. M., Axelrod, B. N., Wall, J. R., Lee, A. J., & Kinikini, D. T. (2015). Is co-norming required? *Archives of Clinical Neuropsychology, 30*(7), 611–633. <https://doi.org/10.1093/arclin/acv039>
- Rury, J. L. (1988). Race, region, and education: An analysis of Black and White scores on the 1917 Army Alpha Intelligence Test. *The Journal of Negro Education, 57*(1), 51–65. <https://doi.org/10.2307/2295276>
- Sayegh, P., Vivian, D., Heller, M. B., Kirk, S., & Kelly, K. (2023). Racial, cultural, and social injustice in psychological assessment: A brief review, call to action, and resources to help reduce inequities and harm. *Training and Education in Professional Psychology*. Advance online publication. <https://www.doi.org/10.1037/tep0000451>
- Scarr, S. (1994). Burt, Cyril L. In R. J. Sternberg (Ed.), *Encyclopedia of intelligence* (Vol. 1, pp. 231–234). Macmillan.
- Schuster E. (1914). Report of the committee to study and to report on the best practical means of cutting off the defective germ plasm in the American population. I. The scope of the committee's work. II. The legal, legislative and administrative aspects of sterilization. *The Eugenics*

- Review*, 6(3), 247–248. Retrieved June 16, 2023 from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2987070/pdf/eugenrev00367-0055.pdf>
- Shuey, A. M. (1958). *The testing of Negro intelligence*. Bell.
- Stetson, G. R. (1897). Some memory tests of whites and blacks. *Psychological Review*, 4(3), 285–289. <https://doi.org/10.1037/h0069814>
- Terman, L. (1916). *The measurement of intelligence*. Houghton Mifflin. <https://doi.org/10.1037/10014-000>
- Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science*, 14(6), 623–628. [https://doi.org/10.1046/j.0956-7976.2003.psci\\_1475.x](https://doi.org/10.1046/j.0956-7976.2003.psci_1475.x)
- Walsh, W. B., & Betz, N. E. (1995). *Tests and assessment*. Prentice-Hall.
- Wang, Z. M. (1993). Psychology in China: A review. *Annual Review of Psychology*, 44, 87–116. <https://doi.org/10.1146/annurev.ps.44.020193.000511>
- Weiss, L. G., & Saklofske, D. H. (2020). Mediators of IQ test score differences across racial and ethnic groups: The case for environmental and social justice. *Personality and Individual Differences*, 161, Article e109962. <https://doi.org/10.1016/j.paid.2020.109962>
- Wiggan, G. (2007). Race, school achievement, and educational inequality: Toward a student-based inquiry perspective. *Review of Educational Research*, 77(3), 310–333. <https://doi.org/10.3102/003465430303947>
- Zhang, H., & Fang, L. (2020). The development of psychological and educational measurement in China. *Chinese/English Journal of Educational Measurement and Evaluation*, 1(1), 56–64. <https://www.ce-jeme.org/journal/vol1/iss1/7>