An Illusion of Equity: The Legacy of Eugenics in Today's Education
By Wendy Zagray Warren

The University Press of Kentucky (August 1, 2023). 270 pp., $40.00. First edition, paperback.
ISBN: 978-0813197685

Reviewed by Benjamin Brumley, West Chester University

Abstract

This review examines Wendy Zagray Warren's An Illusion of Equity: The Legacy of Eugenics in Today's Education through a critical quantitative lens. It considers how standardized testing's historical ties to eugenics continue to shape educational measurement, social sorting, and claims of meritocracy. The review highlights Warren's critique of testing as a mechanism for reproducing racial and class hierarchies, while identifying opportunities for deeper engagement with the history of psychometrics and the consequences of test use. It concludes by proposing two directions for more equitable assessment: participatory, community-accountable approaches and lessons from international alternatives, particularly Finland. Together, these directions position measurement as one source of evidence supporting learning rather than the governing authority of schooling.

Keywords: *critical quantitative research; standardized testing; eugenics; educational equity; psychometrics; assessment*

The use of large scale testing confirms that it has long served as one of [our nation's] caste primary tools. And in a delicate operation at the turn of the twenty-first century, this dehumanizing practice has replaced the beating heart of American public education. -Wendy Zagray Warren, as cited by Eric Jackson.

Wendy Zagray Warren's book *An Illusion of Equity: The Legacy of Eugenics in Today's Education* is a must-read for future educators and educational researchers in assessment. I first encountered this book during Dr. Kimberly Black's presentation at the Critical Questions in Education Conference in Pittsburgh (Black 2024). The ideas in it are provocative and timely for understanding the history of American standardized testing in education. What follows is a critical quantitative review of Warren's book. Drawing on my background in critical quantitative methods in education, I attend both to what Warren argues and to how measurement systems can make certain claims appear natural and scientific, even when those claims may not withstand closer critical scrutiny.

A Critical Quantitative Book Review

I approach this book review through a critical quantitative lens. Critical quantitative research is a relatively new tradition in education that brings together two camps that have often been positioned as opposites in education: quantitative methods (statistics, modeling, measurement, data analysis) and critical theory (the study of power, ideology, history, and structural inequality). Critical quantitative approaches reject the assumption that numbers are neutral reflections of reality (Tabron and Thomas 2023). Instead, these approaches treat data as socially produced: shaped by what institutions choose to count and what is ignored. In short, critical quantitative work asks how numbers come to exist in the forms we encounter and what they do socially once they circulate. These are not new ideas in educational research broadly, but they remain relatively uncommon within quantitative research communities themselves (Tabron and Thomas 2023).

A critical quantitative approach typically treats categories like race, gender, and class as socially constructed and politically maintained (Castillo and Gillborn 2023; Castillo and Strunk 2025). These categories shape where people live, what schools they attend, what resources are available, how they are treated by institutions, and what futures appear plausible. Observed disparities like test-score gaps are therefore interpreted less as evidence of individual deficits and more as symptoms of structural systems that distribute opportunity unevenly. Compared with conventional quantitative research, critical quantitative studies center power and therefore treats fairness as contested rather than settled (see Tabron and Thomas 2023; Castillo and Strunk 2025).

In this review, I apply that lens to Warren's book and to the traditions she draws upon. While Warren's analysis can be read as critical quantitative work, she grounds her framing more explicitly in Isabel Wilkerson's *Caste* (2020) and Ibram X. Kendi's antiracism framework than in the named critical quantitative literature. My goal is to surface the quantitative and measurement stakes of her argument, clarify where her critique is strongest, and suggest where deeper engagement with critical quantitative debates could make her case even sharper. Warren organizes the book into three main sections: (1) "The Illusion and the Spin," (2) "Scaffolding Caste," and (3) "A Culture Framed by Illusions," followed by a brief concluding section, "Hope." I follow that structure here.

Section 1: The Illusion and the Spin

The first section is partly memoir and partly intellectual origin story. Warren recounts her path toward questioning educational testing and foregrounds Stephen Jay Gould's *The Mismeasure of Man* as a core text (Gould 1988). Gould's book remains one of the most influential critiques of intelligence testing and its historical entanglement with race science and eugenics. Even when readers dispute aspects of Gould's argument, his broader contribution is difficult to dismiss: he demonstrates that measurement systems are not merely technical artifacts but cultural projects, projects that can be designed to confirm what powerful people already believe about hierarchy.

From a critical quantitative perspective, Gould's work should be canonical. He shows how easily scientific measurement can become a storytelling machine. When a society begins with a belief in a "natural" ranking of groups, it can build instruments that appear to reveal that ranking. Those instruments then feed back into the belief system: scores become evidence that hierarchy is real, deserved, and inevitable, while institutional arrangements are recast as reflections of natural differences rather than political choices. Warren shares an anecdote about a professor of hers dismissing *The Mismeasure of Man* when she brought it to him as a student. I recognize the scenario

because I have seen versions of it in myself as a scholar. The episode raises a question that Warren's first section implicitly poses but could state more directly: why is a book like Gould's so often treated as peripheral in training for measurement and assessment? If the history of IQ testing and its abuses is foundational to the field's ethical legitimacy, treating that history as elective rather than central signals something about what the field prioritizes.

In Chapter 2, Warren introduces Educational Testing Service (ETS). This is an important pivot because it shifts the critique from the abstract ("tests have problems") to the institutional ("testing is an industry largely run by one large, private, corporation"). Most Americans have taken ETS-linked assessments, or have been shaped by decisions made with them, without ever learning what ETS is, what it does, or how profoundly it has influenced educational pathways. This shift sets up one of Warren's later arguments: that large-scale testing is not merely a measurement practice but a governance strategy. Nicholas Lemann's *The Big Test* (2000) and Allan Nairn's *The Reign of ETS: The Corporation That Makes Up Minds* (1980) are useful companions here because they help readers see standardized testing not only as assessment but as social sorting, an engineered mechanism for selection and legitimacy. Both are valuable background texts that augment Warren's book.

Section 2: Scaffolding Caste

Section 2 develops Warren's core thesis: contemporary schooling is scaffolded by inherited hierarchies, maintained through policy, culture, and, crucially, measurement. Warren draws heavily on Wilkerson's *Caste* (2020) and Kendi's *How to Be an Antiracist* (2019) to interpret how racial hierarchy is reproduced even when institutions adopt the language of equity. I agree with Warren's structural framing. Social class is a core organizing principle of modern testing regimes. In my view, influenced heavily by Nicholas Lemann's *The Big Test* (2000), large-scale standardized testing functions as one of the most powerful class devices in American education.

There is also a frequently untaught history behind U.S. standardized testing: early intelligence tests, the rise of psychometrics, and the development of major college admissions tests were not politically neutral (Brumley, Brumley, and Morrison 2026). They were shaped by Harvard elites such as James Bryant Conant and Henry Chauncey, who sought a scalable sorting mechanism fairer than older forms of aristocratic privilege by repackaging elite selection in the language of standardized intelligence, most notably through the SAT (see Lemann 2000). Once scores become attached to consequential decisions those scores begin to function like a kind of currency. Families with resources can purchase more of it through tutoring and test preparation. Even when individual test items are not explicitly biased, the system can still function as biased because it converts structural advantage into "merit." In this sense, the system can be celebrated as meritocratic opportunity while preserving the deeper structure: some groups have far more access to the conditions that produce high scores, and the system treats that access as irrelevant to what the score supposedly means. For a satirical exposition of this logic, see Michael Young's 1956 *The Rise of the Meritocracy*, which originally framed meritocracy as a dystopian concept.

One piece I find underdeveloped, both in Warren's account and often in U.S.-centered discussions of testing, is the role of non-U.S. statisticians and researchers who helped pioneer many of the foundational methods later used in American psychometrics. British (and British-linked) eugenicists developed key ideas around testing and "normality." Drawing on Adolphe Quetelet's concept of the "average man," they recast the bell curve from a model of measurement error into

a model of human populations, treating the mean as a baseline for what was "normal" and departures from it as "deviations" (Grue and Heiberg 2006). Francis Galton and his proteges then extended this into an explicitly hereditarian political project: they popularized the normal curve for human traits, coined "eugenics," and developed regression and early correlation thinking to quantify how traits appeared to run in families, making population averages and variation look like biological facts with social implications (Zuberi and Bonilla-Silva 2008; Brumley Brumley, and Morrison 2026). Once correlation became a technique for comparing many measures at once, it fed directly into the emerging field of psychometrics: Spearman's *g* was derived from patterns of positive correlations among test scores, and IQ testing later operationalized intelligence by scaling scores onto a normal distribution with a fixed mean and standard deviation. In that way, British eugenics helped shape the cultural meaning of "normal," "abnormal," and "ability" as apparently objective, ranked properties ideas that would later influence the American standardized testing industry (Brumley, Brumley, and Morrison 2026).

Section 3: A Culture Framed by Illusions

Section 3 is the most technical, and arguably the most important, because it moves from historical critique into the internal logic of measurement. Warren explains key concepts in test design and validity and illustrates the self-reinforcing nature of standardized assessment systems. She shows how the technical language of psychometrics can function as a shield: it makes the system appear too specialized to critique, and it reframes political questions as technical questions (Is the test reliable? Are the items aligned?). Warren's central point is that validity is often operationalized in circular ways. A test is deemed valid because it predicts outcomes that are themselves produced by earlier tests or by test-like structures. This argument is not new; it was a central claim in Gould's *The Mismeasure of Man* (1988). New assessments are frequently validated by correlating them with established assessments, which effectively means the new test is rewarded for resembling the old one. The critique matters because validity, in practice, can become institutional continuity disguised as scientific proof. If the system is built around test-based selection, then any test that predicts success within that system will appear valid, even if the system is unjust. A test can be excellent at predicting who will thrive in a stratified environment without that being an ethical justification for the environment itself.

An extension of Warren's argument would emphasize consequential validity: the idea, associated with modern validity theory, that validity concerns not only score meaning but also the consequences of score use. When consequences include life altering tracking, the burden of proof should be higher than predictive correlations alone. One of the most sobering dynamics Warren highlights is testing inertia: testing has become so entrenched that reforms default to more testing, that is better tests, aligned tests, or equitable tests. Tests to solve a testing problem.

Hope: Testing Inertia and Two Directions Forward

Warren's final "Hope" section is brief at only about twelve pages. I found myself wanting more. Still, it is the most generative section because it invites readers to imagine what a liberatory, or critical approach to assessment might look like. A vision that takes measurement seriously through a critical quantitative lens while refusing to let measurement govern the entire meaning of schooling. If we move away from a standardized test-based authority, what replaces it? I offer two directions that others building on Warren's work could expand. These ideas are not new, but they remain viable paths for critical quantitative scholarship.

Recommendation One: Build Assessment from the Ground Up, Not Top Down from the Educational Testing Service

Educational measurement should be driven by more participatory, community-accountable processes. At present, the system is administered by a few large educational testing corporations. In the United States, much of standardized testing is administered by ETS (2025), which is often described as the largest testing corporation in the world. Prior to today's centralized testing culture, teachers more commonly designed authentic performance tasks and portfolios tied to meaningful learning (Taylor 1994; Supovitz and Brennan 1997; Darling-Hammond 1994). Educational communities could instead define what they value, civic reasoning, creative work, collaboration, and multilingual competence, rather than accepting what is easiest to test (Dewey 1916). Assessment would become a tool for learning and reflection on standards rather than a mechanism for ranking and comparison (Taylor 1994). My critique here is anti-authoritarian and anti-corporate: measurement should be one source of evidence among many, not the primary engine of educational governance, and it should be oriented more bottom up rather than determined by one large private corporation.

Recommendation Two: Learn from Alternative Education Models outside the United States, such as Finland

A second direction is internationally comparative: examine alternative education systems that rely less on constant standardized testing, particularly Finland. The United States tests more than many other countries and is more likely to rely on standardized multiple-choice tests (Darling-Hammond 2015, 2017). Finland is a useful counterexample because its education system is widely regarded as one of the strongest in the world, yet it relies relatively little on standardized testing. In Finland, there are no national examinations in comprehensive basic education; assessment is primarily teacher-led, and the major national exam is the matriculation examination at the end of upper secondary schooling. OECD profiles of Finnish evaluation describe a decentralized approach with less top-down control than many other systems (OECD 2024). Finland thus illustrates a structural alternative to U.S. standardized testing: system monitoring and teacher professionalism can substitute for universal high-stakes testing run by a large, private, corporation like ETS.

A Reminder about Eugenics and Standardized Testing

Over the last century, hard lessons about testing have often been learned only after enormous harm. The most unsettling implication of Warren's book is that we may be doing something similar today: continuing to rely on a eugenics-era model of sorting, modernized, rebranded, and wrapped in equity language, while congratulating ourselves for progress. History may not look kindly on how readily we normalized test-driven governance of children. Warren's contribution in *An Illusion of Equity: The Legacy of Eugenics in Today's Education* is to make that issue visible again, reminding us that what appears to be the objective order of the world is often a story about power and who gets to wield it.

References

Au, Wayne. 2022. *Unequal by Design: High-Stakes Testing and the Standardization of Inequality*. New York: Routledge.

Black, K. 2024. "The Legacy of Eugenics in Today's Education: A Book Response to Wendy Warren's *An Illusion of Equity*." Presentation at the Critical Questions in Education Symposium, Pittsburgh, PA, October 28.

Brumley, Benjamin, Lauren Brumley, and Dana Morrison. 2026. "Learning about the Past to Leave It Behind: Eugenics and Norm-Referenced Educational and Psychological Testing." *Critical Questions in Education* 17 (1).

Castillo, Wendy, and David Gillborn. 2023. "How to 'QuantCrit': Practices and Questions for Education Data Researchers and Users." *EdWorkingPaper* 22-546. Annenberg Institute at Brown University. https://doi.org/10.26300/v5kh-dd65.

Castillo, Wendy, and Kamden Strunk. 2025. "How to QuantCrit: Applying Critical Race Theory to Quantitative Data in Education." *All Books and Media by Montclair State University Authors* 608. https://digitalcommons.montclair.edu/all_books/608.

Darling-Hammond, Linda. 1994. "Performance-Based Assessment and Educational Equity." *Harvard Educational Review* 64 (1): 5–31.

Darling-Hammond, Linda. 2015. "Testing, Testing." TEDxStanford video, recorded June 29, 2015.

Darling-Hammond, Linda. 2017. "Teacher Education around the World: What Can We Learn from International Practice?" *European Journal of Teacher Education* 40 (3): 291–309.

Dewey, John. 1916. *Democracy and Education: An Introduction to the Philosophy of Education*. New York: Macmillan.

Educational Testing Service. 2025. "About Us." Accessed March 7, 2026. https://www.ets.org/about.html.

Gillborn, David, Paul Warmington, and Sarah Demack. 2018. "QuantCrit: Education, Policy, 'Big Data' and Principles for a Critical Race Theory of Statistics." *Race Ethnicity and Education* 21 (2): 158–79. https://doi.org/10.1080/13613324.2017.1377417.

Gould, Stephen Jay. 1988. *The Mismeasure of Man*. New York: W. W. Norton & Company.

Grue, Jan, and Anne Heiberg. 2006. "Notes on the History of Normality: Reflections on the Work of Quetelet and Galton." *Scandinavian Journal of Disability Research* 8 (4): 232–46.

Kendi, Ibram X. 2019. *How to Be an Antiracist*. New York: One World.

Lemann, Nicholas. 2000. *The Big Test: The Secret History of the American Meritocracy*. New York: Macmillan.

Nairn, Allan. 1980. *The Reign of ETS: The Corporation That Makes Up Minds*. Washington, DC: Ralph Nader.

OECD. 2024. *Finnish Education Evaluation Centre (FINEEC): OECD Centre for Skills Evaluations*. OECD Skills Studies. Paris: OECD Publishing. https://doi.org/10.1787/ b1c0b194-en.

Supovitz, Jonathan. 2009. "Can High-Stakes Testing Leverage Educational Improvement? Prospects from the Last Decade of Testing and Accountability Reform." *Journal of Educational Change* 10 (2): 211–27.

Supovitz, Jonathan, and Robert Brennan. 1997. "Mirror, Mirror on the Wall, Which Is the Fairest Test of All? An Examination of the Equitability of Portfolio Assessment Relative to Standardized Tests." *Harvard Educational Review* 67 (3): 472–507.

Tabron, Lolita, and Amanda Thomas. 2023. "Deeper Than Wordplay: A Systematic Review of Critical Quantitative Approaches in Education Research (2007–2021)." *Review of Educational Research*.

Taylor, Catherine. 1994. "Assessment for Measurement or Standards: The Peril and Promise of Large-Scale Assessment Reform." *American Educational Research Journal* 31 (2): 231–62.

Wilkerson, Isabel. 2020. *Caste: The Origins of Our Discontents*. New York: Random House.

Zuberi, Tukufu, and Eduardo Bonilla-Silva, eds. 2008. *White Logic, White Methods: Racism and Methodology*. Lanham, MD: Rowman & Littlefield Publishers.

Benjamin P. Brumley is an Assistant Professor in the Department of Educational Foundations & Policy Studies at West Chester University (WCU). He earned a B.A. from the University of North Carolina at Chapel Hill and a Ph.D. from the University of Pennsylvania. His work focuses on educational assessment and measurement, particularly identifying and addressing biases in assessment methods, advocating for child-centered, culturally responsive, and equitable practices. He also researches the historical intersections of eugenics, standardized testing, and how those affect marginalized children.