



Test Scores are Only a Symptom: The Challenge of Seeking to Close Educational Gaps While Ignoring Historical Legacies

Lashawn Richburg-Hayes

Test Scores are Only a Symptom: The Challenge of Seeking to Close Educational Gaps While Ignoring Historical Legacies

Lashawn Richburg-Hayes, January 2021
William T. Grant Foundation

Author's Note:

I am especially grateful to Dr. Vincent Guilamo-Ramos, who told me that my project's struggles were not management-related (as I had thought), but reflective of institutional racism. This feedback started my deeper journey of learning how to understand and address institutional racism as an economist conducting a rigorous evaluation. It made me confront the fear I had in even saying institutional racism, as well as the larger fear of attempting to document it.

Suggested Citation:

Richburg-Hayes, L. (2021). *Test scores are only a symptom: The challenge of seeking to close educational gaps while ignoring historical legacies*. William T. Grant Foundation.

In the quest to address income inequality and intergenerational poverty, it is critical to seek answers to the questions “What works?” and “For whom?” Education is a prime area in which to seek solutions, as one’s level of education completed is positively correlated with income later in life. This connection contributes to a specific focus on schools to reduce inequality and a reliance on test scores as both a measure of intervention effectiveness and a proximal measure of young people’s future outcomes.

As Richard Murnane points out in a new essay, however, some efforts to improve and expand instruction have yielded improvements in test scores, but a narrow focus on this measure alone obscures a larger picture (2021). Murnane goes on to detail three interventions—including two housing mobility interventions and a school choice program—that did not improve students’ reading and math scores but did dramatically improve their life chances. These findings suggest what many educators and parents know: Larger contexts affect school performance. In the United States, for instance, racism is pervasive and continues to have implications for African Americans in all areas of their lives, including health (Kwate et al., 2003; William & Williams-Morris, 2000), income (Rothstein, 2018), and wealth (Akbar et al., 2019). To address income inequality and intergenerational poverty, then, education is just one instrument—we need a tool chest.

Until researchers and funders recognize and attempt to account for the broader context surrounding the studies we conduct, our work will continue to shine a narrow light on potential solutions. Both research and data reflect the perspectives of the researcher and the data collector, as well as the societal context. Given the widespread fallacy that research and data are objective, this recognition will require intentional effort on the part of researchers. Ignoring these facts will only perpetuate recommendations for incomplete policy actions.

Theories of Change Account for Individual Action, but not Racism

Poverty is associated with poor outcomes in many areas of life (Lin & Harris, 2009). While there is evidence that many spells of poverty are relatively brief (Bane & Ellwood, 1986; Stevens, 1999), long spells make up the vast majority for those in poverty at any single point in time. And African Americans have a higher incidence of long spells than other groups (Hoynes et al., 2006).

Poverty and racism are synergistic. Here, racism is defined as the organized system that leads to subjugating certain groups relative to others (William & Williams-Morris, 2000). That is, the definition of racism that I'm employing is synonymous with a racialized system—or structural racism—which does not require a psychological phenomenon of racist beliefs held by individuals (Bonilla-Silva, 1997). In a racialized system, the concepts of economics, politics, societal expectations, and ideology are all partially structured by the placement of people in racial categories. This placement may be reinforced by policy that reflects beliefs of inherent deficits or other internal shortcomings. This view of racism as structural is a critical distinction because it dramatically shifts the theory of change of interventions that seek to improve outcomes among BIPOC with low-income from the individual to the larger context.

For example, the theory of change of Moving to Opportunity (MTO), one of the studies cited by Murnane, is that living in high poverty areas diminishes the life chances of low-income residents (Goering et al., 1999), but vouchers to help families move to lower poverty areas can improve familial outcomes related to income, health, and education. While not based on MTO's specific theory of change, the Chicago Housing Authority's demolition of high-rise public housing buildings beginning in the 1990s represented a natural experiment that could be used to test its validity. Similarly, the theory of change underlying the Charlotte-Mecklenberg school choice voucher program involves a relocation of students from neighborhood schools (assumed to be low-performing, underfunded, and located in poverty-affected neighborhoods) to better educational opportunities (Deming, 2011). While the theory of change in all instances involves a movement to secure better resources, it does not account for the larger system that contributes to the inequities in the first place. In this way, it implicitly assumes that, with the move, individuals have the means to improve their children's outcomes (and in the case of the housing interventions, also improve their economic outcomes as adults).

Racism is Pervasive

It is easy (and perhaps comforting) to regard having low income as a result of individual behaviors and decisions. This narrows the focus of attention to helping the individual make better choices. However, what happens when the individual's choices are limited in numerous ways (subtle and not so subtle) that are outside of the control of most? The result is often that the larger society continues to emphasize the Horatio Alger ideal of pulling one's self up despite one's *personal circumstances*. While there are many BIPOC who have come from extremely challenging circumstances to better their life chances and those of their families, overlooking history and maintaining silence toward barriers to success seems like an additional assault on BIPOC and a reinforcement of a deficit perspective. This is particularly true for African Americans.

For example, even after the abolition of slavery, African Americans have faced structural racism. This has played out, in part, in housing segregation—a nationwide phenomenon that could only be possible with the cooperation of the real estate industry, banking institutions, federal housing policy, and support for neighborhood covenants (Massey & Denton, 1993). Supporting public schools through funding formulas that rely on property taxes only reinforces institutionalized racism as unequal funding models give way to separate and unequal education. Add to this the history of employment limitations, such as historic union policies to limit membership by race and to use African Americans only as strikebreakers (Massey & Denton, 1993), coupled with the movement of high-wage jobs from urban industrial centers to the suburbs, high unemployment rates, and mass incarceration of African American males (contributing to single, female-headed households), and it's not surprising that test scores between African American children in low-income schools lag behind those of their White counterparts. In fact, to compare test scores without acknowledging the costs that are born unevenly by African Americans who make decisions within the context of being Black in America (which can result in constrained *personal decision making*) seems to suggest that individual or family decisions are the cause of disparate outcomes.

The unevenness of historical policy, however, is a causal mechanism that is often unacknowledged (Rothstein, 2018). Differences in outcomes reflect these circumstances, as well as individual decision making (Mullainathan & Shafir, 2013). As the popular YouTube video "[Life of Privilege Explained in a \\$100 Race](#)" suggests (Youngsterdam Dynamo, 2019), it is a fallacy to believe that everyone begins at the same starting point and has an equal opportunity for success when African Americans experience multiple disadvantages relative to their White peers.

While all of this history may seem tangential to the economics of education, these contextual factors make it clear that education is just one tool to address inequality and intergenerational poverty. In the context of historical (and current) barriers, education may be a meager tool for *large-scale* societal change (Berliner, 2013). It is in this context that a reduction in test scores gaps needs to be viewed. Without acknowledging that the gaps themselves are symptomatic of a larger system (Billings-Ladson, 2006), intervention development and policymaking will continue to operate with the false hope of finding a silver bullet solution.

Addressing Root Causes Matters

The three interventions that Murnane highlights are noteworthy precisely because they address housing segregation, a root cause of educational differences (Massey & Denton, 1993; Clark, 1965). While addressing this issue is promising, it is not enough. Is the policy implication to “move folks out of these terrible neighborhoods”? This is ineffective given present-day housing discrimination (Christensen & Timmins, 2019), racial fear tactics (Karni, Haberman, & Ember, 2020; Gomer & Petrella, 2017), and White Americans’ growing concern of reverse discrimination (Earl & Hodson, 2019). But it is also problematic because it presumes that nothing works in high-poverty neighborhoods—a deficit view that focuses on lack. More fruitful avenues may include recommending mixed-income housing for all new developments, policies to ensure that gentrification does not result in displacement, or a dismantling of housing policies that have disparate impact. However, the acknowledgement of history is again needed to prevent the implementation of policies that may devolve into other forms of racism. A poignant cautionary tale is the integration management programs of the 1980s—policies geared to reduce racial segregation through ensuring a minority of Blacks in areas through quotas—which were used to enforce racism in a different manner. Under these programs, units set aside for Blacks typically had long wait lists and were smaller in number, while those for Whites were readily available (Lake & Winslow, 1981, cited in Massey & Denton, 1993).

Policy recommendations can be made more actionable by taking a culturally responsive and equitable evaluation lens to the research in order to determine additional solutions given the evidence around the mechanism (Thomas et al., 2018; Stanfield, 1999). An evaluation is culturally responsive if it considers the behavior, values, customs, common beliefs, and context of participants and the program being evaluated (Hood, Hopson, & Kirkhart, 2020; Frierson et al., 2010). Equitable evaluation encompasses the awareness that linguistic, historical, and socioeconomic differences, along with structural factors, perpetuate inequity in America (Public Policy Associates, 2015). Evaluators can start to address these factors through developing research questions and protocols to account for these factors and use the information to push deeper understanding.

Solutions that address root causes—not just symptoms—will require an interdisciplinary approach that seeks to understand and address multiple areas that contribute to the problem of racial disparities (Billings-Ladson, 2006). This is difficult work, especially given that our most rigorous methodology for assessing causality—random assignment—is not helpful in discerning the impact of a system. Consider that in the MTO study, there is strong internal validity about the outcomes on children associated with the moves to lower poverty neighborhoods, but the history of racialized systems in each of the cities is contextual. The Chicago Housing Authority, for example, is infamous for the *Gautreaux* class action lawsuit that charged the CHA and the U.S. Department of Housing and Urban Development with discrimination in the location of federal housing and the assignment of tenants. The case took 15 years for resolution through a verdict by the Supreme Court and yet Chicago remains one of the most racially segregated cities in America. This context is important to understand the outcomes of MTO.

This Work is Hard, but it is the Right Thing to do

Sadly, none of these points are new. Kenneth Clark addressed them extensively in the 1950s and 1960s, and evaluators have been making the call for more nuanced evaluation efforts for more than 40 years (see, e.g., Thomas et al., 2018; Thomas & Parsons, 2017; Gadson-Billings, 2006; Hood, 2001; Massey & Denton, 1993; and Clark, 1965). Perhaps we have reached a critical juncture in the U.S. where we are able to recognize the injustice that befalls some Americans and we are able to see the benefits our system bestows on others. Given the consequences of reinforcing the status quo, this growing awareness must encourage us to pursue our evaluation work differently.

I am aware that what I am recommending cannot be done within one disciplinary approach. “How” and “why” questions will require an interdisciplinary examination. I also acknowledge that taking this approach may be risky for the careers of some researchers who may not have the support of their colleagues, peers, or institutions. Further, I recognize that some funders—including large federal contracting agencies—will not want to support work that delves into context which seems out of scope.

There are no clear and easy solutions, but there are starting points. Researchers and evaluators alike can begin in four broad areas:

Familiarization with the history and context of the issue(s) studied

There is a history behind every policy and program and, given American history, there is a high likelihood that the history reflects the majority in power (Clark, 1965). As a result, it becomes necessary to understand whether the policy reinforces negative, stereotypic views (House, 2017) and to learn about the perspectives of those affected by policies and programs who lack the power of policy creation. That is, rather than take an agnostic position of looking at an issue in only its current framing, learn about what has been done before (and why) from multiple perspectives, paying special attention to unintended consequences (Richburg-Hayes, 2014) as well as long-term outcomes (House, 2017). This supplemental knowledge may temper naïve policy recommendations and mitigate implicit blame of the victim when context matters (Ross & Nisbett, 2011). Maria Cancian, immediate past president of the Association for Public Policy Analysis and Management, illustrated the value of this approach in understanding child support policy. While research has noted disparate child support remittance rates between Black and White fathers, the research almost uniformly ignored that wage garnishment and other reinforcement policies were based on a White, middle class belief of an absent father's lack of desire to contribute to his children's needs. This underlying tenet is largely not applicable to low-income fathers, who do not have the ability to pay (Cancian, 2020).

Broaden your perspective through personal reflection

Time for personal reflection is critical to processing information that may be new to your worldview or foreign to your discipline. After considering the history and context, take some time to reflect on its meaning and your reactions (especially visceral reactions), and ponder the implications for your research. Reflect on how the information affects your personal beliefs about people and contexts, as well as the truths that you hold (Ganly, 2018). This is not a linear process, but a potentially complex undertaking that may nevertheless help you identify creative additions to your work and inform the recommendations that you make. While I've spent the last year reflecting deeply on race and cultural responsiveness in evaluation, I've come to the realization that I've always thought about it, but I've separated it from my work as it did not seem to have a place in rigorous economic research. I've now reached the point where I understand that the perspective I had—fostered by my graduate training and professional constraints—does not serve the issues that I care deeply about and hope to change through my contributions.

Change an aspect (or aspects) of your research approach.

There are a number of guides to implementing a culturally responsive and equitable approach to evaluation (see Public Policy Associates, n.d.; Hood, Hopson, & Kirkhart, 2015; Frierson et al., 2010). While many of

these guides are focused on evaluation, the tenets apply to primary and secondary research as well. Culturally responsive evaluation can be considered at each stage in the research process (e.g., *We All Count's* Data Equity Framework):

- *Development of research questions:* You can critically consider implicit researcher bias, hidden assumptions, etc., that are embedded in the way that research questions are framed. You can also consider what is *not* being asked and whether this omission is indicative of an overlooked area (Thomas & Parsons, 2017). Finally, you can explore whose perspective is being uplifted/supported/forwarded by the research questions. Stakeholders have different motivations and interests, which do not always align (We All Count, n.d.). Since research questions are reflective of certain stakeholder perspectives, it is important to be explicit about the perspective they represent and consider whether the questions reflect power or influence. The review of the historical literature and understanding of context above can help guide the development of questions and investigation of what is overlooked.
- *Funding:* Consider seeking multiple funders for the work, as some funders are focused only on rigorous quantitative analyses while others are willing to support qualitative research that supplements that approach.
- *Methodological approach:* Most methodologies have limitations (Cody, 2020) that can nevertheless be overcome by augmenting the work with additional approaches. For example, RCTs can be accompanied by implementation research and rich qualitative research to secure perspectives from participants. In general, we can set up research projects that do not stop at zero impact or statistically significant findings, but instead go further by employing the voices and experiences of those directly affected to inform the answer to “why?” Our approach can also incorporate a greater understanding of the context, which may illuminate issues beyond the spotlight topic (e.g., how housing policy can affect educational outcomes). This may provide more nuanced and accessible policy recommendations, which may be more scalable and suffer fewer unintended consequences.
- *Data collection:* Instruments and protocols for data collection need to reflect the stakeholders of interest and be absent of deficit language and stereotypes. Let’s strive to obtain multiple data sources (including qualitative) and examine multiple truths (augmenting causal findings).
- *Analysis:* Critically consider how variables are developed and what they could mean for all audiences. For example, when you have the choice to characterize a binary variable in the positive or the negative, why have you made that choice? Reporting on the proportion of African American

male students that receive a grade of D or lower has a different meaning than a report of those who received a grade of B or higher. This is the choice of the analyst (and may seem like a trivial choice for an outcome variable), but it could have profound implications for whether a program is interpreted as working or not.

- *Reporting:* Researchers have a standard way of reporting that may reinforce deficit framing and paternalistic stances. Consider using a culturally responsive or equitable evaluation checklist to review any publications prior to finalizing. If possible, share your report or major findings with stakeholders reflective of the target population to obtain other perspectives on framing and interpretation.
- *Dissemination:* Go beyond publishing in your field's journal of choice and create engaging, policy-relevant one-pagers, blog posts, infographics, or other materials that help make your findings more accessible to a broader set of stakeholders. Think about developing a strategy in the early stages of your research, which will help you identify stakeholders and share information according to their needs.

Take personal and professional responsibility for change

Finally, we can all take personal and professional responsibility for indicating blind spots in our research and the research of our colleagues (Thomas et al., 2018). We can do this by understanding that there is a system within which we are operating. *We can also lose the fallacy that research is objective and data are unbiased*—they both reflect the beliefs of the researcher, who has a perspective and determines what, how, and why to examine a particular set of data. We can become more courageous in naming racism clearly when it exists or is a factor, rather than omitting this fact or using coded language about social policy (Quadango, 1994). Finally, we can give grace when our colleagues do not quite get it right.

Conclusion

Again, research to reduce inequality rests on questions of “What works?” and “For whom?” But we also need to ask “How?” and “Why?” We need to account for the larger historical and social contexts in which we operate. This point is clear to many BIPOC who are subjects or participants in research: The odds are not fair or even or in your favor. As researchers, seeking to improve these odds while eliding historical legacies will produce only incomplete responses.

While I have suggested that we adopt changes in the ways we conduct research, I am not proposing that this will be a cure-all, or that I know all there is to know about culturally responsive and equitable evaluation. In fact, I am still learning, reflecting, applying, and reiterating in my own work. But while I do not have all of the answers, I have learned a lot more about policy and my power for change as an evaluator than I would have learned without deep reflection on race and equity. I truly believe that acknowledging and encompassing a larger view can help us move forward.

References

- Akbar, P., Li, S., Shertzer, A., & Walsh, R. (2019). *Racial segregation in housing markets and the erosion of Black wealth*. (NBER Working Paper No. 25805). National Bureau of Economic Research. <https://doi.org/10.3386/w25805>
- Berliner, D. (2013). Effects of inequality and poverty vs. teachers and schooling on America's youth. *Teachers College Record*, 115(12), 1–26. <http://www.tcrecord.org/Content.asp?ContentId=16889>
- Bonilla-Silva, E. (1997). Rethinking racism: Toward a structural interpretation. *American Sociological Review*, 62(3), 465–480.
- Cancian, M. (2020). *Family matters: Reflections on who counts and why* [Conference presentation]. Presidential address at the 2020 conference of the Association for Public Policy Analysis and Management. <https://mccourt.georgetown.edu/news/dean-cancians-2020-appam-presidential-address/>
- Chetty, R., Hendren, N., & Katz, L. (2016). The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment. *American Economic Review*, 106(4), 855–902. <http://dx.doi.org/10.1257/aer.20150572>
- Christensen, P., & Timmins, C. (2019). *Sorting or steering: Experimental evidence on the economic effects of housing discrimination* (NBER Working Paper No. 24826). National Bureau of Economic Research. <https://doi.org/10.3386/w24826>
- Clark, K. B. (1965). *Dark Ghetto: Dilemmas of social power*. Wesleyan University Press.
- Cody, S. (2020). *How even the best evidence can yield bad decisions, and what we can do about it* (Working Paper). Insight Policy Research. <https://insightpolicyresearch.com/wp-content/uploads/2020/05/Cody-EBP-Working-Paper-20200916.pdf>
- Deming, D. (2011). Better schools, less crime? *The Quarterly Journal of Economics*, 126(4), 2063–2115. <https://doi.org/10.1093/qje/qjr036>
- Earle, M. & Hodson, G. (2019). Questioning white losses and anti-white discrimination in the United States. *Nature Human Behaviour*, 4, 160–168. <https://doi.org/10.1038/s41562-019-0777-1>
- Frierson, H., Hood, S., Hughes, G., & Thomas, V. (2010). A guide to conducting culturally responsive evaluations. In Frechtling, J. (Ed.), *The 2010 user-friendly handbook for project evaluation* (75–96). National Science Foundation.
- Ganly, T. (2018). Taking time to pause: Engaging with a gift of reflective practice. *Innovations in Education and Teaching International*, 55(6), 713–723. <https://doi.org/10.1080/14703297.2017.1294492>
- Goering, J., Kraft, J., Feins, D., McInnis, D., Holin, M. J., & Elhassan, H. (1999). *Moving to Opportunity for Fair Housing Demonstration Program: Current status and initial findings*. US Department of Housing and Urban Development. <https://www.huduser.gov/portal/Publications/pdf/mto.pdf>
- Gomer, J., & Petrella, C. (2017, October 10). How the Reagan administration stoked fears of anti-white racism. *The Washington Post*. <https://www.washingtonpost.com/news/made-by-history/wp/2017/10/10/how-the-reagan-administration-stoked-fears-of-anti-white-racism/>
- Hood, S., Hopson, R., & Kirkhart, K. (2015). Culturally responsive evaluation: Theory, practice, and future implications. In K. Newcomer, H. Hatry, & J. Wholey. (Eds.), *Handbook of practical program evaluation* (4th edition). Hoboken, NJ: Wiley.
- Karni, A., Haberman, M., & Ember, S. (2020, November 6). Trump plays on racist fears of terrorized suburbs to court White voters. *New York Times*. <https://www.nytimes.com/2020/07/29/us/politics/trump-suburbs-housing-white-voters.html>

- Kwate, N., Valdimarsdottir, H., Guevarra, J., & Bovbjerg, D. (2003). Experiences of racist events are associated with negative health consequences for African American women. *Journal of the National Medical Association, 95*(6), 450–460. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2594553/>
- Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in U.S. schools. *Educational Researcher, 35*(7), 3–12. <https://doi.org/10.3102/0013189X035007003>
- Lin, A., & Harris, D. (2009). *The colors of poverty: Why racial & ethnic disparities persist* (Policy Brief No. 16). National Poverty Center. http://www.npc.umich.edu/publications/policy_briefs/brief16/PolicyBrief16.pdf.
- Mullainathan, S., & Shafir, E. (2013). *Scarcity: Why having too little means so much*. Henry Holt and Company.
- Massey, D. S., & Denton, N. A. (1993). *American apartheid: Segregation and the making of the underclass*. Harvard University Press.
- Murnane, R.J. (2021). *Can policy interventions reduce inequality? Looking beyond test scores for evidence*. William T. Grant Foundation.
- Public Policy Associates (2015). *Considerations for conducting evaluation using a culturally responsive and racial equity lens*. Public Policy Associates. <https://publicpolicy.com/wp-content/uploads/2017/04/PPA-Culturally-Responsive-Lens.pdf>
- Quadango, J. (1994). *The color of welfare: How racism undermined the war on poverty*. Oxford University Press.
- Richburg-Hayes, L. (2014). Incentivizing success: Lessons from experimenting with incentive-based grants. In A. Kelly & S. Goldrick-Rab (Eds.), *Reinventing financial aid*. Harvard Education Press.
- Ross, L. and Nisbett, R. (2011). *The person and the situation: Perspectives of social psychology*. Pinter & Martin.
- Rothstein, R. (2018). *The color of law: A forgotten history of how our government segregated America*. Liveright.
- Stanfield II, J. (1999). Slipping through the front door: Relevant social scientific evaluation in the people of color century. *American Journal of Evaluation, 20*(3), 415–431.
- Thomas, V. and Parsons, B. (2017). Culturally response evaluation meets systems-oriented evaluation. *American Journal of Evaluation, 38*(1), 7–28. <https://doi.org/10.1177/1098214016644069>
- Thomas, V., Madison, A., Rockcliffe, F., DeLaine, K., & Lowe, S. (2018). Racism, social programming, and evaluation: Where do we go from here? *American Journal of Evaluation, 39*(4), 514–526. <https://doi.org/10.1177/1098214018772910>
- We All Count (n.d.) *The data equality framework* [Web page]. <https://weallcount.com/the-data-process/>.
- Williams, D. R., & Williams-Morris, R. (2000). Racism and mental health: The African American experience. *Ethnicity & Health, 5*(3-4), 243–268. <https://doi.org/10.1080/713667453>
- Youngsterdam Dynamo. (2019, July 5). Life of privilege explained in a \$100 race [Video]. YouTube. <https://youtu.be/kyl4EJhq47A>



60 E. 42nd Street, 43rd Floor, New York, NY 10165
wtgrantfoundation.org | info@wtgrantfdn.org | 212.752.0182