

RESEARCH ARTICLE

# “The Magic of Numbers is Strong”: *Hobson v Hansen* and Contested Social Science in Judicial Decision Making

Keith McNamara 

University of Wisconsin-Madison, Madison, WI, USA  
Emails: [kmcnamara3@wisc.edu](mailto:kmcnamara3@wisc.edu); [kthmacnam@yahoo.com](mailto:kthmacnam@yahoo.com)

(Received 14 August 2021; revised 2 December 2022; accepted 30 March 2023)

## Abstract

*Hobson v. Hansen* (1967) is best known as the first federal court case to rule against discriminatory use of standardized tests in the context of educational tracking. It was also significant as one of the first desegregation cases after *Brown v Board of Education* (1954) to use psychological evidence in its ruling. This essay briefly examines the debates over ability testing before *Hobson*, the contexts of post-desegregation D.C. educational politics that shaped the case, the social scientific evidence presented in the case, and its application to the court's ruling. It argues that while scholars have correctly acknowledged the court's mistaken assumptions about testing, the evidence presented of testing bias nevertheless cogently illustrated a broader constellation of discriminatory District practices. A review of the testimony suggests that while the psychological evidence was central to the court's ruling, the opinion rested less on the resolution of social scientific debates over testing bias than it did on the need to determine the justification of ability testing in the context of District tracking practices. Although sweeping in scope, the decision did little to resolve long running disputes over ability testing. Instead, it only helped inaugurate a more heated and contentious legal environment for educational testing in the coming decades.

**Keywords:** Courts; education; psychology; testing; bias

## Introduction

The year 1967 was a landmark in the history of educational testing. In a sweeping, passionately written decision, Judge James Skelly Wright of the U.S. District Court for the District of Columbia declared a series of district school policies – the track system, optional school zones, teacher assignment policy, and unequal school funding – unconstitutional on grounds of both *de jure* and *de facto* segregation (Bickel 1967). Handed down more than ten years after *Brown v. Board of Education* (1954), *Hobson v. Hansen* (1967) represented another bold judicial intervention in school policy. It also came during a period of rising civil rights agitation among the

© The Author(s), 2023. Published by Cambridge University Press on behalf of the Social Science History Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

city's African American population and rapidly deflating public confidence over the condition of the city's public schools. By the 1950s, the growing influence of achievement tests and the overall increase in scores had produced "a revolution of rising expectations" (Diner 1990: 124). But by the late 1960s declining test scores, growing complaints about disciplinary problems, and mounting criticisms from Congress and the press dissolved public confidence in its schools (Diner 1990: 124–127).

While not as influential as *Brown, Hobson* was nevertheless significant as the first federal case to rule on the legality of aptitude and achievement tests in the context of educational tracking (Jensen 1980: 28; Note 1973: 1042). It was the first case to challenge the use of group ability tests as the basis for placing children into special education classes and was the first time a court had used the presence of disproportionate numbers of black students in low-ability classes as evidence of bias (Thorndike 2005). After *Hobson*, courts became more receptive to questioning similar testing practices in the context of language minorities (*Diana v. Board of Education* 1970) as well as in the context of individually administered ability tests (*Larry P. v. Riles*, 1972, 1979), effectively inaugurating court battles over educational testing that continued throughout the 1970s and 1980s.

Like *Brown*, the plaintiffs relied on a substantial amount of social scientific evidence to make their case (Rossell 1980: 244). Historian John Hogan cites *Hobson* (along with *Brown*) as the exception to Federal courts' proclivity for deciding educational cases based on "legal precedent alone without support from the findings of studies in education and psychology" (Hogan 1970: 289). Although the social scientific evidence in *Brown* has enjoyed substantial scholarly attention (e.g., Heise 2005; Jackson 2001, 2005; Mody 2002), the social science evidence in *Hobson* has received far less scrutiny (exceptions include Bersoff 1979; Kirp 1973; Rossell 1980; Shea 1977). In general, this scholarship has been critical of the court's interpretation of the psychometric evidence presented in the case.

David Kirp argues that the *Hobson* court made an "analytic misstep" by equating ability with innate characteristics. He contends that by stressing the discriminatory effects of testing rather than the educational deprivation of the tracking system, the court's abolition of the tracking system was "unresponsive to the problem at hand" (Kirp 1973: 767–768). Similarly, Christine Rossell asserts the court "misunderstood" the issue in part due to the plaintiffs' experts' misleading testimony about "the ability of tests to accurately measure the innate intelligence" of children from disadvantaged backgrounds. Citing only one of the plaintiffs' studies, Rossell claims the opinion "reads like an exercise in illogic" (Rossell 1980: 275–276). Donald Bersoff likewise describes the court's "gravest error" from a psychometric standpoint was its insistence that "grouping can only be based on tests that measure innate ability" (Bersoff 1979: 50–51).

These scholars correctly admonish the court's mistaken assumption that tests should measure "innate ability," a claim few credible psychologists would have made. But if the court admitted and considered substantial social science evidence about the validity and potential biases of ability testing, its conclusions were not simply misunderstandings, missteps, or illogical errors. A review of the testimony suggests that while the psychological evidence was central to the court's ruling, the opinion rested less on the resolution of social scientific debates over testing bias than

it did on the need to determine the justification of ability testing in the context of District tracking practices.

This essay briefly examines the debates over ability testing before *Hobson*, the contexts of post-desegregation D.C. educational politics that shaped the case, the social scientific evidence presented in the case, and its application to the court's ruling. I will argue that while scholars have correctly acknowledged the court's mistaken assumptions about testing, the psychological evidence of testing bias nevertheless cogently illustrated a broader constellation of discriminatory District practices. Although sweeping in scope, the decision did little to resolve long running disputes over ability testing. Instead, it only helped inaugurate a more heated and contentious legal environment for educational testing in the coming decades.

### The historical context of ability testing

The first major public debates over ability testing emerged after World War I when the rapid deployment of intelligence tests in schools generated controversy within and outside the psychological profession (Brown 1992; Chapman 1993; Cravens 1988; Thomas 1982, 1984). Their widespread adoption between 1890 and 1930 coincided temporally with their promised ability to solve many burgeoning social, economic, and political problems. The increasing cultural authority of science (especially quantitative science), the dramatic growth of secondary school enrollments and student diversity, and the pressures felt by early American psychologists to establish themselves as a legitimate scientific profession like other natural and social science disciplines also encouraged testing (Brown 1992; Camfield 1973; Fass 1980, 1991; Samelson 1979; Tyack 1974).

By the 1930s, school leaders adopted cost efficient measures of student classification and placement to respond to swelling secondary school enrollments and widespread curtailment of public school budgets. Both psychologists and the public accepted the credibility and utility of IQ measures as intelligence testing became deeply institutionalized in schools across the country (Carson 2006: 266–270; Chapman 1993: 128–145; Fass 1980). Despite the ideological drift within the social sciences toward environmental explanations and a shifting emphasis within psychology toward measuring abilities beyond IQ, testing critics found themselves increasingly marginalized (Brown 1992: 138–139; Chapman 1993).

By 1945, most psychologists abandoned the prospect of accurately distinguishing between innate and learned ability. However, postwar expansion of higher education enrollments promoted the proliferation of ability tests, and they encountered only limited public scrutiny before the late 1950s (Ackerman 1995: 291–297; Carson 2006: 258–270; Capshew 1999). By then, ability tests became widely associated with identifying the academically “gifted,” receiving renewed federal support through the passage of the National Defense Education Act in 1958. But opponents of *Brown* and desegregation used black-white test score differences to argue that integration would have disastrous consequences (Ackerman 1995; Jackson 2005; Porter 2017).

The *Brown* decision prompted powerful white resistance from pro-segregationist southerners. The most fervent opponents pursued strategies including public school

closures, “freedom of choice” policies, “pupil placement” laws, intimidation, and overt violence to avoid desegregation orders (Klarman 1994; Patterson 2001: 86–117; Note 1962). By 1966, the U.S. Department of Health, Education, and Welfare adopted stronger enforcement guidelines by threatening to sever federal educational funding from segregated school districts. In response, southern white officials began to segregate *within* desegregated schools. Many instituted tracking systems to segregate black and white students into separate classes based on standardized test scores (Dickens 1996: 472–473; Klarman 1994: 84; Note 1989: 1323; Patterson 2001: 139–140).

Meanwhile, pro-segregationist social scientists began a concerted effort to challenge the *Brown* Court’s supposed erroneous findings of fact. They attacked the social scientific arguments in *Brown* that segregation caused psychological harm to black students, that race could not be a rational basis for segregation, and that there were no significant differences between the races in terms of learning ability (Allport et al. 1953). They collectively marshalled psychological, anthropological, and sociological evidence to prove that the disparate black-white test score gaps used to defend segregation were scientifically valid measures of biological and immutable racial differences (Garrett 1947; Shuey 1958). Their goal was to crystallize a scientifically “objective” challenge that would undermine the *Brown* Court’s supposedly flawed assumptions about racial equality and segregation’s pernicious psychological consequences (Jackson 2005: 118–131).

However, challenges to *Brown*, though successful in the sympathetic federal courts of the South, were subsequently reversed by the U.S. Circuit Courts as contrary to the Supreme Court’s ruling that segregated schooling was “inherently unequal” (*Stell v. Savannah-Chatham Board of Education* 1964; *Evers v. Jackson Municipal Separate School District* 1964, 1966). Nevertheless, if *Brown* had settled the issue over whether *racial* classifications could be legitimately used to segregate students, it left *ability* classification *prima facie* constitutionally permissible.<sup>1</sup> Many northern districts embraced classifications based on *individual* ability, spurred by Cold War concerns about cultivating the talents of the “gifted” (Porter 2017, 2018). However, in the contexts of post-*Brown* (and *Bolling*) Washington, D.C., the permissibility of ability testing and tracking faced strong opposition in the shifting climate of opinion by the late 1960s.<sup>2</sup>

### The local context of *Hobson*: education in the district after *Brown* and *Bolling*

Before 1954, the public schools in the District had been racially segregated since the first law was passed in 1862 to provide primary schools to African American

<sup>1</sup>“(T)here is no constitutional prohibition against an assignment of individual students to particular schools on a basis of intelligence, achievement or other aptitudes upon a uniformly administered program,” however, “race must not be a factor in making the assignments” (*Stell v. Savannah-Chatham County Board of Education* 1964: 62).

<sup>2</sup>*Bolling v. Sharpe* (1954) was the companion case to *Brown* that applied specifically to the District. The Court had to rule separately in *Bolling* since the Fourteenth Amendment applied only to states. The Court maintained that “it would be unthinkable that the same Constitution would impose a lesser duty on the Federal Government” (*Bolling v. Sharpe* 1954: 347).

children (Roe 2004/2005). Between the end of World War II and 1954, total school enrollments grew beyond the capacity of the existing school system. Although the demographic changes through black migration and white flight resembled that of other large northern cities, the District was unique in its historic role as a sanctuary for freed slaves and as a long established center of the black intelligentsia (Asch and Musgrove 2017; Mintz 1989; Moore 1999). It was also unique as one of the earliest black majority cities in the nation (Diner 1991: 91). Those demographics made educational politics in the District after *Bolling* especially volatile. Although by 1966 African Americans made up about 68 percent of the population, their enrollment in the public schools was over 90 percent (ibid.; Richards 2004/2005: 25).

Resistance among white residents to desegregation was endemic between 1945 and 1954. After *Bolling*, locally organized opposition collapsed, and little of the violent “massive resistance” that appeared in other cities occurred in the District following the decision (84<sup>th</sup> U.S. Congress 1956; Clement 2004/2005: 99–104). However, challenges to desegregation continued, mostly from southerners in Congress. In 1956, Congressman James Davis of the House District Committee scheduled a hearing composed of almost all southern legislators (and signatories to the “Southern Manifesto”) to publicly vilify the District’s desegregation efforts. The hearing, entitled “To Investigate Public School Standards and Delinquency in the District of Columbia,” was designed to publicize the supposedly ruinous consequences of desegregation. Assistant superintendent of DC Schools Carl Francis Hansen defended the District’s desegregation efforts as “a miracle of social adjustment” despite withering criticism from southern Congressmen (Hansen 1957). Hansen would soon be admired as a tenacious advocate of *de jure* desegregation (Clement 2004/2005: 99–104). The 1956 Committee would not be Hansen’s last defense of school policies before a hostile audience. But it would be the beginning of a decade long career in the hot seat of District school politics in the wake of *Brown* and *Bolling*.

### Carl Hansen and the four track system

Although many in the press looked upon Washington D.C. as a “Model for the Nation” in carrying out the Supreme Court’s desegregation orders, rising expectations among black residents clashed with worry among white residents over declining academic standards (Diner 1990: 120–122). Public concern over standards in the District schools was nothing new, but the rapid desegregation efforts after 1954 caused considerable public agitation from segregationists and civil rights activists alike. They cited evidence that district-wide standardized testing revealed that average scores of African American students were lower than the national average. By Hansen’s own account, even supporters of desegregation voiced these concerns (Hansen 1964: 11–13).

While skeptics and opponents of desegregation seized on the test score data to argue their case, Hansen insisted that the differences had nothing to do with race or integration, but with the dramatically unequal educational opportunities perpetuated by segregation (Hansen 1964: 11). His handling of desegregation in the District won acclaim in the first few years after his appointment as

Superintendent in 1958. Born and educated in Nebraska, Hansen's career as a teacher, school principal, and assistant superintendent helped to shape his view that every student should have an academically rigorous education. He earned a reputation as a no-nonsense administrator committed to a back-to-basics curriculum to navigate the challenges of integration and raise overall academic performance in the District. He was touted as a "passionate believer in meliorism in education," as DC Board of Commissioners President Walter Tobriner described him, "[with] the courage to translate his beliefs into practical school programs" (Koerner 1961). Although committed to integration, Hansen was also aware of the difficulties stemming from white flight and the concentration of underprepared African American students in greater numbers in the District schools (Hansen 1964: 28–31).

But Hansen also believed that the main responsibility of the public schools was to "promote intelligent behavior" through a traditional liberal arts curriculum available to all students regardless of socioeconomic background or prior academic preparation (Hansen 1960b: 126–127). His Four Track Curriculum thus emphasized academic skills appropriate to different levels of scholastic ability. Each track would have separate eligibility requirements and curricular expectations at the elementary, junior high, and senior high levels. For example, at the high school level, most students were assigned to either the *regular* (*college prep*), for "the college able pupil not qualified or not wanting to take the more demanding honors curriculum," or the *general* curriculum, for "the pupil unqualified for or not electing the honors or college preparatory curricula" (Hansen 1964: vi–vii, 131–132). A minority of students could qualify at the high end into the honors track, for "the exceptionally able pupil," while students placed into the lowest *basic* track received a sequence "required of the severely academically retarded high school student."<sup>3</sup>

Initially implemented in the high schools in 1956, the Four Track Curriculum was designed to wed Hansen's ideals of universal education with the local realities of vastly different academic starting points. By 1959, the system had been expanded to the junior high and elementary levels as well (Hobson 1978: 8). Designed for targeted instruction at exceptionally high and low ability levels, it aimed to provide foundational academic content for all students. By strategically serving the needs of a wide range of students, it embraced the democratic ideals of the comprehensive high school (Hansen 1964: v–x). Hansen claimed the track system supplied "maximum challenge for the gifted as well as the less gifted," and by 1960 argued that both white and black students "enjoyed educational conditions . . . superior to those available under the previous policy of racial segregation in the schools" (Hansen 1960b: 216, 223). Praised by the Board of Education and widely supported by both white and black residents, Hansen gained a national reputation as one of the most effective urban school leaders in the tumultuous early days of desegregation.<sup>4</sup>

---

<sup>3</sup>*Ibid*; Hansen's use of the word "retarded" implied academic underachievement that had "complex origins," and involved "organic, functional, and cultural factors," citing President Kennedy's Panel on Mental Retardation as a source. He argued that because of this complexity, "the methods of working with it educationally must also be sophisticated, precise, and multidisciplinary."

<sup>4</sup>One of the main African American newspapers, for example, stated how it took "pride in saluting Dr. Hansen as the new school superintendent." (The Afro-American 1958).

## Growing dissatisfaction

The public honeymoon would not last long. What became for Hansen the Achilles heel of his system stemmed from the problems of the lowest track, the so-called “basic” track, which critics argued disproportionately channeled low-income black students into dead end trajectories. To Hansen, the basic track was a more humane alternative to “shunting” students into a “sidetrack of so-called nonacademic education,” and aimed to prevent dropouts by more meaningfully adapting to the needs of lower performing – and presumably lower ability – students (Hansen 1960a: 125–126). Hansen believed a basic track in the comprehensive high school could resolve many of the criticisms leveled at separate schools for the “academically retarded”: more efficient use of resources, equitable access to quality teachers, better community and parent cooperation, lower likelihood of negative labeling, exposure to more advanced students, and the possibility for students to advance to more challenging curricular levels when ready (Hansen 1964: 37–43). But the reality of how the track system operated was starkly at odds with Hansen’s original vision. Public complaints accumulated by the early 1960s that both the school board and eventually Congress were compelled to address. Responding to criticisms about the basic track, the board sponsored a special study in 1964 by the Urban League to review district policy and make recommendations. Pressure intensified after the board ignored the study’s findings and recommendations that outlined racial discrimination in the districts’ policies (Baratz 1975: 68).

Complaints continued over the next year and a half from mostly black civic and religious leaders. The criticisms were many: the basic track was a “dumping ground”; students were never able to learn enough to “test out” to a higher track; students were stigmatized and suffered low morale from placement in the basic track. Critics highlighted the failure of the basic track to provide students with the knowledge and skills needed for college admissions. And there was growing skepticism about the accuracy of the tests used to place students in the basic track. Rising opposition among parents, civil rights groups, and labor and religious organizations eventually caught the attention of Congress, which held a series of hearings on the District schools starting in late 1965. The hearings paid especially close attention to complaints about the track system (89<sup>th</sup> U.S. Congress 1966: 35–36).

## Congress steps in: the Pucinski report

The Task Force on Anti-Poverty in the District of Columbia, colloquially known as the Pucinski Committee after its Chairman, Roman C. Pucinski (D-IL), focused its attention on evaluating “the degree to which the public school system of the Federal City has been neglected.” The Committee cited “charges brought to its attention that widespread discrimination exists in the public schools in the Nation’s Capital and that some of the programs . . . not only help perpetuate segregation, but . . . help ‘freeze’ youngsters into a future of poverty” (89<sup>th</sup> U.S. Congress 1966: 1). Among the many features examined in its Report, the track system received particularly heavy criticism. It indicated a “strong suspicion” that students tested as “dull” might actually prove capable of going to college through more effective testing programs.

Despite expressing profound respect for Hansen, it recommended either revising the whole track system or “dropping it entirely” (89<sup>th</sup> U.S. Congress 1966: 3).

Most damning to Hansen’s credibility as a meticulous administrator, the Report revealed that only in the fall of 1965 did individual testing become mandatory for children recommended for placement in the basic track. It noted that prior to the policy change, any student scoring below 75 on an IQ test or 3 years below grade level on an achievement test was automatically assigned to basic. It further cited a “crash testing” of students in September 1965 after the discovery of large numbers of students being placed in the basic track by principals without the requisite psychological testing. When finally evaluated, only 441 out of the 653 elementary and 620 junior high students tested accurately as belonging to the basic track. Worse, in his testimony to the Task Force, Hansen admitted that “more refined techniques are being used (in Washington) to determine whether children are mentally retarded or have average ability but perform at the level of mentally retarded because of background and other factors.” This suggested that the District’s testing program was inherently flawed, leading to the improper placement of pupils (89<sup>th</sup> U.S. Congress 1966: 36–43). Though the Task Force recognized Hansen’s efforts to remedy the many problems besetting the school system, its damaging findings gave significant fodder for opponents of the tracking system. And it became a potent weapon in the hands of civil rights activists waiting to deliver a decisive blow against the District (89<sup>th</sup> U.S. Congress 1966: 2).

### The outrage of Julius Hobson

That opportunity came even before the Pucinski Report was published. One of those civil rights activists, a statistician for the Social Security Administration, sued in federal court on behalf of his daughter who was placed into the basic track. Julius Hobson had by then established a long record as a civil rights agitator and District gadfly, beginning with his first local fight as a PTA member through his leadership in the NAACP where he initiated a suit against the Metropolitan Police Department in 1957 alleging racism in police officer promotions. In 1961, he became president of the Congress of Racial Equality (CORE), leading its membership in a series of dramatic public protests, including picketing and boycotting retail stores that would hire only white employees, staging “live-ins” at private buildings that excluded black renters, and leading a 4,500-person march to protest unfair housing policy (Franklin 1977; Gorney 1977).

Born in Birmingham, Alabama, Hobson as a child worked at a library where he cleaned the floors but was not allowed to take out books. He graduated from the only high school in the city that admitted black children and was a decorated World War II veteran by the time he enrolled at Columbia University and Howard University for graduate work in economics. Though skilled as a researcher, his more publicly provocative persona became even more useful to his work as a civil rights activist. His style often wed public confrontation with a taste of the theatrical. During a sit-in at the Washington Hospital Center, Hobson snuck up to an all-white ward, climbed onto one of the beds and declared that he would move only if arrested. The Hospital obliged but was thereafter compelled to desegregate its wards

once the press got wind of Hobson's publicity stunt. And in 1964, Hobson drove through the affluent Georgetown neighborhood in a pickup truck with a cageful of possum-sized rats, threatening to unleash the vermin if the city continued to ignore the burgeoning rat problem east of the Park.

Hobson's activism emerged during a period of heightened agitation around issues of home rule, housing, and employment discrimination in which violent white reactions to peaceful civil protests by organizations like SNCC, SCLS, and CORE spurred a "dramatic new phase" of confrontational politics. Hobson, the most "strident of a new generation of black leaders," helped usher in that growing militancy. In comparison to most cities in the United States in the early 1960s, Washington D.C. boasted the nation's wealthiest and most educated black community, and civil rights activists had dismantled many of the Jim Crow laws that their southern counterparts continued to struggle against. Moreover, President Johnson supported D.C. home rule as part of his broader civil rights initiatives and threw his weight behind a 1965 bill that would have created an elected city council. These early signs of progress led to rising expectations that the black community would finally "reap the full fruits of freedom," including good jobs, decent homes, and equal educational opportunities. But like so many before it, the home rule bill passed the Senate only to die in the House. Along with entrenched racial inequalities, including rising unemployment, stagnant wages, deteriorated neighborhoods, and overcrowded schools, many saw Congress' persistent resistance to home rule as affirmation of an unjust and racist system (Asch and Musgrove 2017, 332–347; Diner 1991, 91–92, 95–96).

As in his federal lawsuits against Hansen and the school board, Hobson's unique brand of public pugilism was complemented by his considerable research skills to document and demonstrate the discriminatory practices he frequently railed against (Franklin 1977; Gorney 1977). Hobson's diligence as a litigant easily matched his bravado as a civil rights provocateur. Yet his statistical acumen would not always pay dividends in federal court. Despite the groundbreaking use of psychological testimony in the *Brown* case, few courts were receptive to using social scientific evidence to decide cases about fundamental constitutional questions (Yudof 1978). Luckily for Hobson, the judge assigned to the case was receptive to such evidence, and his legal team took full advantage of it.

### A judge to be reckoned with

The D.C. Circuit Court judge assigned to hear the case was no ordinary federal judge. Judge James Skelly Wright, appointed to the District of Columbia Circuit by President John F. Kennedy in 1962, had built a national reputation for judicial diligence, efficiency, and above all, considerably liberal sympathies for the plight of disadvantaged claimants.<sup>5</sup> Wright had earned both plaudits and condemnations for his unflinching enforcement of desegregation in the face of widespread public opposition while sitting as a federal district judge in New Orleans (Bernick 1980:

<sup>5</sup>Although Wright sat on the D.C. Circuit Court (the appellate Court above the D.C. District Court), he was assigned to the D.C. District (trial) Court bench specifically, and temporarily, for the *Hobson* case (Harvard Law Review Association 1968).

979–992; Siedman 2015: 75–78). Enduring threatening letters, phone calls, protests, and social and professional ostracism, Wright developed an affinity for socially marginalized communities and cultivated a strong judicial commitment to redressing social and economic inequities (Siedman 2015: 75). Even the token desegregation of four black elementary school girls earned him public denunciations as “Judas Scalawag Wright,” his image burned in effigy, petitions for his impeachment, and the need for police to be posted at his home for his personal protection (Bernick 1980: 971–972, 986–990). Despite a professional reputation for efficient and sober adjudication, his record for taking liberal positions on several issues involving racial discrimination and desegregation worried the defendants’ counsel enough to prompt the defendants to unsuccessfully petition to reverse the *Hobson* decision based on Wright’s “bias or prejudice” in the case (Motion to Remand 1967).

On June 19, 1967, Judge Wright handed down his opinion. Citing the precedent of the Supreme Court’s ruling in *Bolling v. Sharpe* (1954), Wright stated the “basic question presented” to the court was “whether the defendants, the Superintendent of Schools and the members of the Board of Education . . . unconstitutionally deprive the Districts Negro and poor public school children of their right to equal educational opportunity with the District’s white and more affluent children.” Ruling on behalf of the plaintiffs, Wright articulated several “findings of fact,” including: the aptitude tests used to assign students to tracks were standardized primarily on white middle class children; the tests didn’t relate to Negro or disadvantaged children; they inappropriately relegated students to lower “blue collar” tracks that denied them educational and occupational opportunities, created a stigma, and reduced expectations to a degree that seldom allowed students to escape. His decree ordered the school district to abolish the track system, to terminate optional zones that allowed white students to escape integrated schools, to transport students from overcrowded schools to underpopulated ones, and to produce plans for equalization in pupil assignment and the integration of the faculty (*Hobson v. Hansen* 1967: 406–408). In 118 pages, the opinion reflected painstaking deliberation over a mass of psychological and social science evidence.

### The psychological evidence: Sources of consensus and disagreement

That evidence was considerable. Although many expert witnesses opined about the track system, the bulk of the opinion’s attention to testing’s role in it rested on the testimony of three psychologists: Dr. Roger Lennon and Dr. John Dailey for the defendants, and Dr. Martin Cline for the plaintiffs. All three agreed that the tests, particularly those labeled “intelligence” or “aptitude” tests, did not – and could not – accurately assess “innate” ability independent from environmental influences. They also agreed that environmental influences, including socioeconomic status, parental education, and various indicators of cultural exposure, and “deprivation” largely accounted for the gap in average test scores between black and white students (Conant 1964; Reissman 1962).

Although these areas of agreement reflected consensus within the psychological profession and the testing community (APA 1966: 10), more noteworthy were the

areas of disagreement. Expert witnesses disagreed about the nature of aptitude tests and their educational value. They disagreed about whether intelligence and achievement tests really measured different constructs or whether their labels were just “the use of two separate words or expressions covering in fact the same basic situation” (Coleman and Cureton 1954: 347; Kelley 1927: 64). They differed over whether the tests were appropriate for use with nonwhite and non-middle-class students. And they offered markedly different opinions about the likelihood that careless interpretation of scores would harm minority children. These disagreements were evident in their testimony about three issues: the accuracy of aptitude tests to predict future performance, the extent that environmental influences rendered test scores inaccurate when applied to minority students, and the degree to which the standardization process and the establishment of national norms rendered traditional tests biased against the majority black student population of the District.

On the defense side, Dr. Roger Lennon, research director for one of the nation’s leading test publishers, staunchly vouched for the validity of the most common aptitude and achievement tests on the market. Acknowledging the error of earlier generations of psychologists who interpreted test scores as measures of fixed and hereditary mental traits, Lennon noted the change in professional opinion in the wake of accumulated evidence about environmental influences (Lennon ND: 12–15). He argued that although standardized tests were “imperfect instruments,” they were far better than the subjective and biased judgment of teachers, and as “independent yardsticks” provided unbiased information about student aptitudes and achievement “attainable in no other fashion” (ibid. 43–44).

Dr. John Dailey, the defendant’s other psychological expert, echoed many of these sentiments. A George Washington University professor with several years of experience on the research team for Project Talent, Dailey’s testimony similarly acknowledged the widely differing environments of low income and minority children while simultaneously defending traditional tests as appropriate evaluative instruments. Though admitting his own studies with low-income black children in the District showed improvements in test scores when questions were given in a more familiar “dialect,” he nevertheless defended the applicability of standard test measures to evaluate the academic competencies of black and white children alike (*Hobson v. Hansen* Transcripts 1967: 6279–6285).

On the other side was the testimony of plaintiffs’ expert, Dr. Marvin Cline, a social psychologist at Howard University Medical School’s Institute for Youth Studies. Dr. Cline offered the most potent rebuttal against the notion that these tests produced valid measures for disadvantaged minority children. While agreeing with Dailey’s assertions that aptitude tests provided scant evidence of “innate” potential, he went much further in denying they could reliably predict anything meaningful about low income or minority kids. For Cline, even the most well established individually administered tests failed to adequately predict how a child might do in the future, and as such, would likely produce harm when used with certain vulnerable populations (ibid.: 1316–1318). Cline conceded that there was indeed extensive literature to demonstrate strong correlations between aptitude and achievement scores, as Lennon suggested. But he insisted that such correlations, and

thus the predictive validity of the tests, only applied to white middle-class children (ibid.:1401–1405).

Much of the disagreement over whether tests were valid measures for low income and African American children stemmed from their disagreements over the degree that environmental factors contributed to the meaningful interpretation of test scores. All three experts agreed about the voluminous research indicating close correlations between socioeconomic status, parental income level, and test scores. But they differed in their assessment of the weight that should be given to the growing body of research on factors like test anxiety, linguistic differences, and teacher expectations (ibid.: 1361–1380, 1401–1409, 1729–1779; Lennon ND: 41–42, 104–105). Both Dr. Lennon and Dr. Dailey acknowledged that cultural and linguistic differences could readily influence test scores, but both still believed that intelligence, aptitude, and achievement tests were credible and valid instruments for predicting future academic achievement (*Hobson v. Hansen* Transcripts 1967: 6287–6290, 6361–6375). In contrast, Dr. Cline argued that while cultural and linguistic disadvantages began in the home, it was the school environment that tended to have the most powerful effect, positive or negative, on student achievement. In Cline’s view, even segregation itself in the context of high concentrations of low-income black children in the District schools could depress test scores and undermine efforts to accurately assess student aptitudes (ibid.:1729–1734, 1378–1380, 1401–1403).

The most technically complex evidence concerned the standardization process in test construction and the development of norms. The most common accusation, both in the plaintiffs’ expert testimony and in the critical literature cited in the opinion, was the charge that these tests were standardized on a white, middle-class population, and therefore biased against populations with dissimilar socioeconomic and cultural backgrounds (ibid.: 1362; See also Goslin 1967: 6–11; Eells et al. 1951; Sexton 1961). Dr. Cline argued that all standardized tests were limited in the range of abilities they could accurately discern, and that they were ill equipped to measure individual characteristics outside the middle range of the population it was normed to (*Hobson v. Hansen* Transcripts 1967: 1362). He contended this was true whether the standardization population was middle class white children, low-income black children, or any other sample population (ibid.: 1363).

Dr. Lennon disagreed that the most common tests were *solely* standardized on white middle-class populations. Intimately familiar with the principles of norming and test construction as the research director of a major test publisher, Lennon suggested the ideal practice when establishing norms was to draw the standardization population from a wide range of geographic, socioeconomic, and cultural groupings (Lennon ND: 23). Yet despite these assurances, he was unable to testify to what extent low-income black children were likely to be represented in the standardization population for many of the tests used in the District (ibid.: 6–8). He also conceded that most nationally normed standardized tests drew mainly from predominantly white middle-class populations for their standardization samples (ibid.: 71).

Therein was the problem. Because the District of Columbia’s population was so unlike a nationally representative standardization sample, critics like Dr. Cline argued these nationally normed tests were inaccurate with respect to the majority of

the District's students (*Hobson v. Hansen* Transcripts 1967: 1712). He suggested that only through the development of *local norms* and the construction of tests standardized on the local population could the tests be appropriate for most students in the District schools (ibid.: 1724). Local norming was not a standard practice in school districts, and it had not yet appeared in the recommendations and ethical guidelines for testing by the American Psychological Association (APA 1950, 1966). All three expert witnesses considered it an important *option* for testing departments in certain circumstances. However, the testimony of Dr. Lennon and Dr. Dailey, and its omission in the APA's *Standards for Educational and Psychological Tests and Measures*, suggests it was not by any means a *requirement*, ethically or otherwise.<sup>6</sup>

Each of these issues – the predictive validity of ability tests, the environmental factors that influenced test scores, and the potential bias against minorities on account of the standardization process – had all been sources of debate within psychology and psychometric testing for decades. Disputes about differences between “intelligence” and “achievement” tests began as far back as the late 1920s, when psychometrically trained psychologists demonstrated substantial overlap (90 to 95 percent) between the two constructs (Anastasi 1984: 129–140; Kelley 1927). Socioeconomic and environmental influences on test scores had been noted as early as the first Binet–Simon tests developed in the early 1900s and became central to a burgeoning social science literature from the 1920s onward (Binet and Simon 1916: 316–321; Richards 2012: 139–181). Even the idea of using “local norms” to assess special populations more fairly was an idea first proposed as early as 1915 by Robert M. Yerkes, the Harvard psychologist most famously known for his role in leading the testing program of U.S. Army recruits during World War I (Yerkes and Anderson 1915).

Thus, the evidence presented in *Hobson* was well within the mainstream of debates in psychology by the late 1960s. The sources of consensus among the experts were consistent with current professional opinion, and their points of departure had been contested issues within psychology since the first decades of the twentieth century. However, any indication that the evidence was incomplete, uncertain, or subject to decades of debate did not register in the confident, often strident, tone of the opinion. Sympathetic to the struggles of desegregation advocates in the wake of southern resistance to *Brown*, Judge Wright was keenly receptive to the plaintiffs' evidence highlighting the discriminatory effects of the tracking system and the testing practices that supported it.

### Wright's weighing of the evidence

Court transcripts reveal that Judge Wright admitted a wide range of expert testimony to inform his decision. The defense had attempted to argue that, while their own experts testified that the tests used in the District could not accurately assess innate potential, their use for sorting and tracking students according to

<sup>6</sup>Though advocacy for local norms recommended in a paper drafted by a working group of the Society for the Psychological Study of Social Issues (SPSSI) suggests a shift in opinion by some members of the profession (Fishman et al 1964).

perceived academic ability was nevertheless justified. For example, John Dailey, when pressed by plaintiffs' counsel, admitted "intelligence tests" did not measure "potential abilities" but rather "developed abilities," including "all the learning opportunity" a student had from home, neighborhood, and school. Despite many potential influences on a student's performance at any given time, "you have to estimate what he will be able to do in the future" (*Hobson v. Hansen* Transcripts 1967: 6291–6293). In defending their use in the District schools, Dailey later added that while it would be "very unfair if you were to say naively, this is a test that measures innate ability and this shows how stupid this kid is because he can't do something," he did not "think it unfair to measure the lack of development that occurred for some children if the purpose for measuring that is to assist in development" (ibid.: 6360–6362). What defendants' experts failed to properly account for was not whether the tests could accurately *predict* success in the future, given their *present environmental circumstances*, but whether the tracking system provided the kinds of compensatory supports to *overcome* those circumstances.

Though the transcripts indicate Wright took the defendants experts' testimony seriously, the opinion shows he was clearly more persuaded by the plaintiffs' testimony. He agreed with Dr. Cline's assessment on the research about teacher expectations leading to a "self-fulfilling prophecy" of low expectations for students placed in the basic track (*Hobson v. Hansen* 1967: 484). He was persuaded by Dr. Cline's (and the SPSSI's) recommendations regarding local norms, while scornful of the District's failure to consider those recommendations (ibid.: 487–488). Importantly, he concurred with plaintiffs' arguments that since "the aptitude tests used to assign children to the various tracks are standardized primarily on white middle class children," they did not "relate to the Negro and disadvantaged child." Consequently, the track assignments based on such tests relegated black and disadvantaged children to an inferior education "from which, because of the reduced curricula and the absence of adequate remedial and compensatory education, as well as continued inappropriate testing, the chance of escape is remote."

All these conclusions accurately reflected the evidence presented by plaintiffs of the track system's rigidity, and the discriminatory practice of placing low performing children in inadequate compensatory education programs from which mobility to higher tracks was unlikely. But Wright also attacked aptitude testing on grounds that none of the psychological experts on either side suggested was warranted. In his assessment of the testimony of the use of ability tests, Wright concluded there was "substantial evidence that defendants presently lack the techniques and the facilities for ascertaining the innate learning abilities of a majority of District schoolchildren." Without such techniques and facilities, he opined, the defendants could not "justify the placement and retention of these children in lower tracks on the supposition that they could do no better, given the opportunity to do so" (ibid.: 488). Indeed, *none* of the experts suggested tests could accurately ascertain the innate learning abilities of students, though it *was* clear from the evidence of the basic track that its lowered expectations and curricular standards had not improved educational opportunities.

Weighing the defense experts' cautions about careful interpretation of test scores, Wright averred that "for many students, interpretation cannot provide meaningful information." He cited "ample evidence . . . that for disadvantaged children group

apptitude tests are inappropriate for obtaining accurate information about innate abilities.” Since defendants had not explained how “interpretation can overcome these technical limitations on the tests,” he found that for most District school children there was “a substantial risk” of being wrongly labelled as having “subnormal intelligence,” a label that could not be effectively removed “simply by interpreting aptitude test scores” (ibid: 489).

Wright’s frequent references to the failure of tests to discern “innate ability” certainly echoed many of the criticisms of ability tests since the 1920s and 1930s (Chapman 1993: 128–145; Franklin 1980; Pastore 1978). Nevertheless, even as *explicitly* hereditarian interpretations of test scores became increasingly marginalized in educational and psychological discourses by the 1930s, *implicit* hereditarian assumptions among educators, administrators, and guidance counselors about individual differences in native ability persisted (Porter 2020). While Wright’s opinion displayed a naiveté about the consensus of professional psychological opinion by the late 1960s, it nevertheless expressed a commonly held assumption among educators, school administrators, and policy makers. It also exposed an unresolved tension within psychometrics – the inability of *any* instrument to accurately tease out “innate” factors from “environmental” factors that influenced learning:

“It will be recalled that a scholastic aptitude test is constructed . . . so as to make possible an inference about an individual’s innate ability to succeed in school . . . A crucial assumption . . . is that the individual is fairly comparable with the norming group in terms of environmental background and psychological make-up . . . Because of the impoverished circumstances that characterize the disadvantaged child, it is virtually impossible to tell whether the test score reflects lack of ability - or simply lack of opportunity . . .” (Hobson *v* Hansen 1967: 485).

In the post-*Brown* context of southern deployment of social scientific evidence to justify old tropes about “innate” racial differences, however, that continuing tension within the discipline had potentially perverse consequences.<sup>7</sup> Whether Judge Wright fully understood that none of expert witnesses believed these tests could accurately assess “innate” ability, he was aware of the dangerous assumptions that low test scores could promote:

“There can be no disputing the fact that teachers universally tend to be strongly influenced in their assessment of a child’s potential by his aptitude test scores. Defendants’ own expert, Dr. Lennon, acknowledged this to be the common experience; and it would defy common sense to think the situation could be otherwise. Although test publishers and school administrators may exhort against taking test scores at face value, the magic of numbers is strong . . .” (Hobson *v* Hansen 1967: 488).

<sup>7</sup>For example, many of the scientists who testified on the pro-segregationist side in *Stell* and *Evers* were members of international eugenics organizations and were generously funded for their research on racial differences (Jackson 2005; Tucker 1994, 2002).

More importantly, his judgment about their inappropriate application to tracking low performing African American students highlighted their misuse in the context of a much larger constellation of discriminatory practices. Wright noted the “critical infirmities” of the track system “when tested by the principles of equal protection and due process,” was to “deprive the poor and a majority of the Negro students in the District of Columbia of their constitutional right to equal educational opportunities” (ibid: 512). The track system as operated by the District had become “a system of discrimination founded on socio-economic and racial status rather than ability.” Noting the law had a “special concern for minority groups for whom the judicial branch of government is often the only hope for redressing legitimate grievances,” Wright professed the court would “not treat lightly” evidence that educational opportunities were being allocated “according to a pattern that has unmistakable signs of invidious discrimination.” What the defendants failed to meet was their burden to explain “why the poor and Negro should be those who populate the lower ranks of the track system” (ibid: 514).

In the court’s view, that system was sustained unmistakably by the testing that supported it:

“What emerges as the most important single aspect of the track system is the process by which the school system goes about sorting students into the different tracks. This importance stems from the fact that the fundamental premise of the sorting process is the keystone of the whole track system: That school personnel can with reasonable accuracy ascertain the maximum potential of each student and fix the content and pace of his education accordingly. If the premise proves false, the theory of the track system collapses, and with it any justification for consigning the disadvantaged student to a second best education” (ibid: 473–474).

Thus, the question before the court – whether the District school system unconstitutionally deprived black and poor children of equal educational opportunity compared with the District’s white and affluent children – was indeed a question where social science evidence, particularly psychological evidence about the testing used to track students, was relevant. That Wright based his decision partly on assumptions about tests that most psychologists no longer held made little difference to recognizing their potential misuse.

### **Contrasting cultures of social science and law**

This discrepancy between the scientific evidence presented and Wright’s employment of that evidence illustrates the contrasting epistemological orientations of science (including social science) and law. Historian Tal Golan observes that science and law are “mutually supporting belief systems and deeply connected social institutions” (Golan 2004: 1–2). But, as legal scholar Stephen Goldberg has noted, they are fundamentally distinct cultures (Goldberg 1994). Goldberg argues that scientists generally seek “empirically verifiable truth” often expressed through traditional causal analyses or probabilistic equations. Judges, in contrast, confronted

with “the pressing need to resolve a social dispute peacefully” will often resort to “patchwork solutions” specific to the problems before them (Goldberg 1987: 1349–1350). These contrasting orientations of science and law also diverge in their professional objectives and putative time horizons. As historian Sheila Jasanoff argues, although the cultures of science and law do share common features – both claim authority to evaluate evidence and rationally derive conclusions, and both rely on credible observations and rule governed methods of assessing facts – they differ in their approaches to fact finding. Science is mainly concerned with getting the facts “right” (within existing paradigms), even to the extent of suspending judgment in anticipation of further evidence. Law is equally concerned with establishing facts correctly, but only for purposes of fairly and efficiently settling disputes. Fact-finding in science is provisional and tentative, always open to further revision or even disconfirmation. Fact-finding in law, by contrast, is time bound, and must take a position and seek closure once the evidence, or the time allowed for it, is exhausted (Jasanoff 1995: 9–10).

This view of science, especially social science, as tentative, provisional, open to competing interpretations, and subject to revision and even disconfirmation over time has long been a source of criticism over its use in judicial decision making at least since *Brown* (Cahn 1955; Clark 1959; Dworkin 1977; O’Brien 1980; van den Haag 1960; Weschler 1959). As Goldberg and others have noted, courts are not the best venues to adjudicate complex and competing social scientific claims, and certainly not long running disputes within disciplines like psychology (Goldberg 1987: 1341–1388; Lindman 1989). If debates over ability testing and cultural bias in testing minority students remained unresolved within psychology by the late 1960s, their debut in federal court could hardly have created an ideal venue for resolution. The court applied psychological evidence to assess whether the tracking system of the DC schools violated constitutionally protected equal educational opportunity, not to resolve longstanding tensions within the discipline. What mattered to the court was not whether there was a *consensus* on the issue of testing bias within the psychological profession (there wasn’t). It was whether the evidence of bias was sufficient to proscribe the use of the tests in the context of DC tracking policies. Judge Wright clearly believed that there was.

Moreover, his embrace of the psychological evidence in the opinion was also a function of its relative coherence in explaining discriminatory practices he was already inclined to believe. Unlike the methodologically sophisticated but often convoluted studies that bedeviled other desegregation and school finance cases, the psychological evidence in *Hobson* was relatively accessible to non-scientists like Judge Wright. Nevertheless, although clearly useful in his assessment of whether the testing practices of the District Schools violated the rights of minority students, Wright lamented the need for the court to rely on social scientific evidence to decide on matters of controversial social policy:

“It is regrettable, of course, that in deciding this case this court must act in an area so alien to its expertise. It would be far better indeed for these great social and political problems to be resolved in the political arena by other branches of government. But these are social and political problems which seem at times to defy such resolution. In such situations, under our system, the judiciary must

bear a hand and accept its responsibility to assist in the solution where constitutional rights hang in the balance. So it was in *Brown v. Board of Education*, *Bolling v. Sharpe*, and *Baker v. Carr* . . . So it is here” (*Hobson v. Hansen* 1967: 517).

### Aftermath of the case

Wright’s intent to address social and political problems in the District that “def[ie]d such resolution” in the political arena would soon encounter significant headwinds. The immediate reaction to the ruling was a mix of enthusiasm, skepticism, and outrage.<sup>8</sup> But its effects on local policies and practices were muted. Although the school board quickly implemented some parts of Wright’s decree, they were less unified about how to implement others. The District moved quickly to abolish optional zones, rearrange boundary lines, and integrate faculty more equitably. But while the track system was formally eliminated, there was little evidence that ability grouping was entirely abandoned within the context of individual schools. Without a system of accurate data collection, the school board could only “hope” that invidious testing and placement practices had been discontinued (Cuban 1975: 20–21).

And the hope that the Wright decree would equalize educational opportunities would soon evaporate. Only three years after the ruling, Julius Hobson found himself once again before the court, this time to complain about unfulfilled court mandates to equalize resources and teacher assignments in a case sometimes known as *Hobson II* (*Hobson v. Hansen* 1971; Horowitz 1977: 106–170). While Hobson would spend the better part of the following decade fighting to hold the District accountable for the Wright decree, his court adversary did not enjoy as much career success in the wake of the decision.

For Carl Hansen, efforts to remediate the growing problems in the schools came too little, too late. Commissioned by Hansen in response to growing complaints, a massive district-wide study of the D.C. school system by Professor A. Harry Passow of Columbia University was released the same day as the *Hobson* decision. It blamed many of Hansen’s policies for the deplorable condition of the school system (Passow 1967). Just like the Wright decree and the Pucinski Report, it called for the end of the track system, the bussing of black children to under-enrolled white schools, greater integration of faculty, and an equalization of financial resources (ibid; Cuban 1975: 16–17). Hansen’s policies and leadership of the school system had been criticized by Congress, a federal court, and now his own commissioned study. When the school board refused to appeal the *Hobson* decision, it subsequently ordered the Superintendent not to appeal. Clearly exasperated, Hansen resigned in protest (Hansen 1967). As he told a reporter in 1965, “I can’t understand it. Here, I kept the lid on for years. And now the Negro community is just about saying I’m a racist” (Jacoby 1967).

Hansen’s fall from grace should not be interpreted as anomalous. His professional trajectory from anointed savior of the public schools to public pariah among the African American community neatly parallels the similarly dramatic

<sup>8</sup>For example, letters to Wright ranged from the laudatory to the violently threatening (Wright Papers 1966–1972).

transformation within American liberalism between the hopeful optimism of the immediate post-*Brown* years and the growing disenchantment and fragmentation of the late 1960s (Patterson 1997). However, if Hansen's descent was not anomalous, the court ruling that accelerated his professional demise in many ways was. Indeed, although 1967 was a landmark in the history of educational testing in that a federal court had now determined that testing and tracking practices could potentially violate constitutional protections, it did little to placate controversy over the validity of ability tests in the coming decades.

In fact, controversies would only intensify. In February of that year, psychologist Arthur Jensen delivered a paper to the American Educational Research Association entitled "Social Class, Race and Genetics: Implications for Education," which reignited debates about racial differences in intelligence (Jensen 1968). And at the American Psychological Association's Annual convention the following year, a splinter group of African American psychologists formed the Association of Black Psychologists (ABP), submitting a list of grievances to the parent organization protesting what they deemed a failure to adequately address critical social issues like poverty and racism. Among their demands was an immediate moratorium on "comparative testing and evaluation programs . . . pending the thorough review and reassessment of the issue on the highly questionable validity" of standardized psychological tests (Williams 1974).

Debates that had breached a federal courtroom had now penetrated the conference halls of educational research groups and provoked new cleavages in the psychological profession. If *Hobson* had minor consequences for the everyday practices of the Washington school system, it nevertheless inaugurated a line of legal challenges to ability testing in the following decades (*Diana v. Board of Education* 1970; *Moses v. Washington Parish School Board* 1971; *Covarrubias v. San Diego Unified School District* 1971; *Guadalupe Organization, Inc. v. Tempe School District No. 3* 1971; *Berkelman v. San Francisco Unified School District* 1974; *Larry P. v. Riles* 1972, 1979; *PASE v. Hannon* 1980). These cases all relied to some extent on the precedent set in *Hobson* that exclusive use of test scores to place children in special education, and differential placements based on race, could be unconstitutional. Many of these cases were more consequential than *Hobson*; they also perpetuated debates about testing.<sup>9</sup>

## Conclusion

*Hobson* may have been the "pioneer case" of educational misclassification (Note 1973: 1039), but the sweeping decision by the court was nevertheless only a minor reverberation in a much longer and unresolved dialectic over educational testing. The extensive testimony presented by psychological experts on both sides of the case revealed a stable consensus developed over decades of progress in psychometric testing. But it also revealed substantive disputes within psychology about the extent to which standardized ability tests were appropriate for minority and low-income

<sup>9</sup>For example, both *Larry P. v. Riles* and *PASE v. Hannon* involved nearly identical issues and evidence, and even many of the same experts, yet the judges in each case came to precisely opposite conclusions (Elliott 1987).

students. Those disputes became especially salient in the context of rising disillusionment with the District's desegregation efforts and its failure to mitigate widening racial and socioeconomic disparities in educational opportunity.

Moreover, Wright's ruling that the District's tracking program was discriminatory to minority students, based largely on the application of standardized ability tests, could not have settled longstanding disputes within the psychological profession over the validity of those tests with minority populations. Even though most psychologists had long abandoned the idea that these tests measured "innate ability," the court's ruling suggested their use in assessing the intellectual potential of minority students nevertheless carried the implication that they could indeed do just that. Although *Hobson* did not proscribe the use of ability tests outside the context of the District's rigid tracking program, it nevertheless highlighted the inherent perils of using such tests in the wake of rising civil rights conflicts over racial segregation and historically unequal educational resources and opportunities. Those perils would only multiply in the coming decades.

In the context of post-*Brown* (and *Bolling*) desegregation cases, the *Hobson* decision was the first of many to challenge the legality of testing and tracking practices (Shea 1977; Bersoff 1979; Rossell 1980). In the following decades, courts would likewise wrestle with often unsettled and contradictory social science evidence to resolve contested educational policy questions over desegregation, complex school finance litigation, and school choice (Heise 2008). But if the decision failed to change educational practices much in the short run, its penetrating legal scrutiny of standardized ability testing did even less to settle enduring controversies over testing bias. If *Hobson* was a minor case in the post-*Brown* desegregation era, it was nevertheless a bellwether in signaling a more ominous and contentious landscape for educational testing.

## Archival Sources

- Hansen, Carl F.** (1967) "Statement." July 3, Carl Hansen Bio File, Charles Sumner Archives, Washington, D.C.
- Hobson v. Hansen** (1967) Court Transcripts. Kansas City, MO: National Archives Federal Records Center.
- Jacoby, Susan** (1967) "Hansen Resigns as D.C. School Superintendent." Washington Post July 4, Carl Hansen Bio File, Charles Sumner Archives, Washington, D.C.
- Koerner, James D.** (1961) "Carl F. Hansen of Washington, D.C." Saturday Review 16: 49-51; Carl Hansen Bio File, Charles Sumner Archives, Washington, D.C.
- "Motion to Remand and Reverse with Directions to Vacate the Judgment of June 19, 1967."** Julius Hobson Papers, Box 10, Folder 6, DC Public Library Special Collections.
- The Afro-American** (1958) "Dr. Hansen Helped to Make 'Miracle'" Carl Hansen Bio File, Charles Sumner Archives, Washington D.C.
- Wright, J.S. Papers** (1966–1972) Boxes 89, 90, Library of Congress Manuscript Reading Room, Washington D.C.

## References

- 84th United States Congress** (1956) Investigation of Public School Conditions, Hearings Before the Subcommittee to Investigate Public School Standards and Conditions, and Juvenile Delinquency in the District of Columbia of the Committee on the District of Columbia, House of Representatives. Washington D.C.: Government Printing Office.

- 89th United States Congress** (1966) Investigation of the Schools and Poverty in the District of Columbia: Hearings before the Task Force on Antipoverty in the District of Columbia. Washington D.C.: House Committee on Education and Labor.
- Ackerman, Michael** (1995) "Mental testing and the expansion of educational opportunity." *History of Education Quarterly* 35 (3): 279–300.
- Allport, F. H.** et al. (1953) "The effects of segregation and the consequences of desegregation: A social science statement." *Minnesota Law Review* (37): 427–39.
- American Psychological Association** (1950) "Ethical standards for the distribution of psychological tests and diagnostic aids." *American Psychologist* 5 (11): 620–26.
- (1966) *Standards for Educational and Psychological Tests and Measures*. Washington, D.C.: American Psychological Association.
- Anastasi, Anne** (1984) "Aptitude and achievement tests: The curious case of the indestructible strawperson," in Barbara S. Plake (ed.) *Social and Technical Issues in Testing: Implications for Test Construction and Usage*. Lawrence, KS: Erlbaum Associates: 129–40.
- Asch, Chris Meyers, George Derek Musgrove** (2017) *Chocolate City: A History of Race and Democracy in the Nation's Capital*. Chapel Hill, NC: University of North Carolina Press.
- Baratz, Joan C.** (1975) "Court decisions and educational change: A case history of the D.C. public schools, 1954-1974." *Journal of Law and Education* 4 (1): 63–80.
- Bernick, Michael S.** (1980) "The unusual odyssey of J. Skelly Wright." *Hastings Constitutional Law Quarterly* 7 (4): 971–1000.
- Bersoff, Donald** (1979) "Regarding psychologists testily: Legal regulation of psychological assessment in the public schools." *Maryland Law Review* 39 (1): 27–120.
- Bickel, Alexander** (1967) "Skelly Wright's sweeping decision." *New Republic*, July 8.
- Binet, Alfred and Theodore Simon** (1916) *The Development of Intelligence in Children (the Binet-Simon Scale)*, trans. Elisabeth S. Kite. Nashville: Williams and Wilkins Co.
- Brown, JoAnne** (1992) *The Definition of a Profession: The Authority of Metaphor in the History of Intelligence Testing, 1890-1930*. Princeton: Princeton University Press.
- Brown v. Board of Education** (1954) 347 U.S. 483.
- Bolling v. Sharpe** (1954) 347 U.S. 497.
- Cahn, Edmond** (1955) "Jurisprudence." *New York Law Review* (30): 150–69.
- Camfield, Thomas M.** (1973) "The professionalization of American psychology, 1870-1917." *Journal of the History of the Behavioral Sciences* 9 (1): 66–75.
- Capshew, James** (1999) *Psychologists on the March: Science, Practice, and Professional Identity in America, 1929-1969*. New York: Cambridge University Press.
- Carson, John** (2006) *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republic, 1750-1940*. Princeton, NJ: Princeton University Press.
- Chapman, Paul Davis** (1993) *Schools as Sorters: Lewis M. Terman, Applied Psychology, and the Intelligence Testing Movement, 1890-1930*. New York: New York University Press.
- Clark, Kenneth B.** (1959) "The desegregation cases: criticism of the social scientist's role." *Villanova Law Review* 5 (2): 224–40.
- Clement, Bell** (2004/2005) "Pushback: the white community's dissent from *Bolling*." *Washington History* 16 (2): 86–109.
- Coleman, William, and Edward E. Cureton** (1954) "Intelligence and achievement: the 'jangle fallacy' again." *Educational and Psychological Measurement* (14): 347.
- Conant, James Bryant** (1964) *Slums and Suburbs*. New York: New American Library.
- Cravens, Hamilton** (1988) *The Triumph of Evolution: The Heredity-Environmental Controversy, 1900-1941*. Baltimore: Johns Hopkins University Press.
- Cuban, Larry** (1975) "Hobson v. Hansen: a study in organizational response." *Educational Administration Quarterly* 11 (2): 15–37.
- Dickens, Angela** (1996) "Revisiting *Brown v. Board of Education*: how tracking has resegregated America's public schools." *Columbia Journal of Law and Social Problems* (29): 469–503.
- Diner, Steven** (1990) "Crisis of confidence: public confidence in the schools of the nation's capital in the twentieth century." *Urban Education* 25 (1): 112–37.
- (1991) "From Jim Crow to home rule." *Wilson Quarterly* 13 (1): 90–101.

- Dworkin, Ronald** (1977) "Social sciences and constitutional rights – the consequences of uncertainty." *Journal of Law and Education* (6): 3–12.
- Eells, Kenneth, Allison Davis, Robert Havighurst, Virgil Herreck, and Ralph Tyler** (1951) *Intelligence and Cultural Differences: A Study of Cultural Learning and Problem-Solving*. Chicago: University of Chicago Press.
- Elliott, Rogers** (1987) *Litigating Intelligence: IQ Tests, Special Education, and Social Science in the Courtroom*. Dover, MA: Auburn House Publishing Company.
- Evers v. Jackson Municipal Separate School District** (1964) 232 F Supp 241; (1966) 328 F.2d 208.
- Fass, Paula** (1980) "The IQ: a cultural and historical framework." *American Journal of Education* (88): 431–58.
- (1991) *Outside In: Minorities and the Transformation of American Education*. New York: Oxford University Press.
- Fishman, Joshua A., Martin Deutsch, Leonard Kagan, Robert North, and Martin Whiteman** (1964) "Guidelines for testing minority group children." *Journal of Social Issues* 20 (2): 129–44.
- Franklin, Ben A.** (1977) "Julius W. Hobson, a black activist in Washington for 20 years, dies." *New York Times*, March 24.
- Franklin, Vincent P.** (1980) "Black social scientists and the mental testing movement, 1920-1940," in Reginald Jones (ed.) *Black Psychology*. New York: Harper and Row: 201–15.
- Garrett, Henry E.** (1947) "Negro-white differences in mental ability in the United States." *Scientific Monthly* 65 (4): 329–33.
- Golan, Tal** (2004) *The History of Expert Testimony in England and America*. Cambridge, MA: Harvard University Press.
- Goldberg, Steven** (1987) "Reluctant embrace: law and science in America." *Georgetown Law Journal* 75 (4): 1341–88.
- (1994) *Culture Clash: Law and Science in America*. New York: New York University Press.
- Gorney, Cynthia** (1977) "Julius Hobson Sr. dies." *Washington Post*, March 24.
- Goslin, David A.** (1967) *Criticisms of Standardized Tests and Testing*. New York: College Entrance Examination Board.
- Harvard Law Review Association** (1968) "Hobson v. Hansen: judicial supervision of the color-blind school board." *Harvard Law Review* 81 (7): 1511–27.
- Hansen, Carl F.** (1957) *Miracle of Social Adjustment: Desegregation in the Washington, DC Schools*. New York: Anti-defamation League of B'nai B'rith.
- (1960a) "Ability grouping in the high schools." *Atlantic*, November Issue: 123–27.
- (1960b) "The scholastic performance of Negro and white pupils in the integrated public schools of the District of Columbia." *Harvard Education Review* 30 (3): 216–36.
- (1964) *The Four-Track Curriculum in Today's High Schools*. Englewood Cliffs, NJ: Prentice-Hall.
- Heise, Michael** (2005) "Brown v. Board of Education, footnote 11, and multidisciplinary." *Cornell Law Review* (90): 279–320.
- (2008) "Judicial decision making, social science evidence, and equal educational opportunity: uneasy relations and uncertain futures." *Seattle University Law Review* 31 (4): 863–90.
- Hobson, Julius** (1978) "Educational policy and the courts: the case of Washington, D.C." *Urban Review* 10 (1): 5–19.
- Hobson v. Hansen** (1967) 269 F.Supp. 401; (1971) 327 F.Supp 844.
- Hogan, John C.** (1970) "The role of the courts in certain educational policy formation." *Policy Sciences* 1 (3): 289–97.
- Horowitz, Donald L.** (1977) *The Courts and Social Policy*. Washington, D.C.: Brookings Institution.
- Jackson, John P., Jr.** (2001) *Social Scientists for Social Justice: Making the Case against Segregation*. New York: New York University Press.
- (2005) *Science for Segregation: Race, Law, and the Case against Brown v. Board of Education*. New York: New York University Press.
- Janasoff, Sheila** (1995) *Science at the Bar: Law, Science, and Technology in America*. Cambridge, MA: Harvard University Press.
- Jensen, Arthur R.** (1968) "Social class, race, and genetics: implications for education." *American Educational Research Journal* 5 (1): 1–42
- (1980) *Bias in Mental Testing*. New York: The Free Press.

- Kelley, Truman L.** (1927) *Interpretation of Educational Measurements*. Yonkers, NY: World Book Company.
- Kirp, David** (1973) "Schools as sorters: the constitutional and policy implications of student classification." *University of Pennsylvania Law Review* **121** (4): 705–97.
- Klarman, Michael J.** (1994) "How Brown changed race relations: the backlash thesis." *Journal of American History* **81** (1): 81–118.
- Lennon, Roger T.** (n.d.) *Testimony of Dr. Roger T. Lennon as expert witness on psychological testing in the case of Hobson, and others vs. Hansen, and others (Washington D.C. Schools)*. New York: Harcourt, Brace & World.
- Lindman, Constance R.** (1989) "Sources of judicial distrust of social science evidence: a comparison of social science and jurisprudence." *Indiana Law Journal* **64** (3): 755–68.
- Mintz, Steven** (1989) "A historical ethnography of black Washington, D.C.," in *Records of the Columbia Historical Society* 52. Washington, D.C.: The Historical Society of Washington, D.C.: 235–53.
- Mody, Sanjay** (2002) "Footnote 11 in historical context: Social science and the Supreme Court's quest for legitimacy." *Stanford Law Review* (54): 793–829.
- Moore, Jacqueline M.** (1999) *Leading the Race: The Transformation of the Black Elite in the Nation's Capital, 1880-1920*. Charlottesville, VA: University Press of Virginia.
- Notes** (1962) "The federal courts and integration of southern schools: troubled status of the Pupil Placement Acts." *Columbia Law Review* **62** (8): 1448–79.
- (1973) "The legal implications of cultural bias in the intelligence testing of disadvantaged school children." *Georgetown Law Journal* (61): 1027–1066.
- (1989) "Teaching inequality: the problem of public school tracking." *Harvard Law Review* (102): 1318–41.
- O'Brien, David M.** (1980) "The seduction of the judiciary: social science and the courts." *Judicature* **64** (1): 8–21.
- Passow, Harry A.** (1967) *Toward Creating a Model Urban School System: A Study of the Washington, D.C. Public Schools*. New York: Teachers College, Columbia University.
- Pastore, Nicholas** (1978) "The army intelligence tests and Walter Lippmann." *Journal of the History of the Behavioral Sciences* (14): 316–27.
- Patterson, James T.** (1997) *Grand Expectations: The United States, 1945-1974*. New York: Oxford University Press.
- (2001) *Brown v. Board of Education: A Civil Rights Milestone and Its Troubled Legacy*. New York: Oxford University Press.
- Porter, Jim Wynter** (2017) "A 'precious minority': constructing the 'gifted' and 'academically talented' student in the era of *Brown v. Board of Education* and the National Defense Education Act." *Isis* **108** (3): 581–605.
- (2018) "The entanglement of racism and individualism: the U.S. National Defense Education Act of 1958 and the individualization of 'intelligence' and educational policy." *Multiethnica* (38): 3–17.
- (2020) "Guidance counseling in the mid-twentieth century United States: measurement, grouping, and the making of the intelligent self." *History of Science* **58** (2): 191–215.
- Reissman, Frank** (1962) *The Culturally Deprived Child*. New York: Harper Row.
- Richards, Graham** (2012) *Race, Racism, and Psychology: Towards A Reflexive History*, 2nd ed. London, UK: Routledge.
- Richards, Mark David** (2004/2005) "Public school governance in the District of Columbia: a timeline." *Washington History* **16** (2): 23–25.
- Roe, Donald** (2004/2005) "The dual school system in the District of Columbia, 1862-1954: origins, problems, protests." *Washington History* **16** (2): 26–43.
- Rossell, Christine H.** (1980) "Social science research in educational equity cases: a critical review." *Review of Research in Education* (8): 237–95.
- Samelson, Franz** (1979) "Putting psychology on the map: ideology and intelligence testing," in A. R. Buss (ed.) *Psychology in Social Context*. New York: Irvington Publishers: 103–68.
- Sexton, Patricia Cayo** (1961) *Education and Income: Inequalities of Opportunity in our Public Schools*. New York: Viking Press.
- Shea, Thomas E.** (1977) "An educational perspective of the legality of intelligence testing and ability grouping." *Journal of Law & Education* **6** (2): 137–58.

- Shuey, Audrey** (1958) *The Testing of Negro Intelligence*. New York: Social Science Press.
- Siedman, Louis Michael** (2015) "J. Skelly Wright and the limits of legal liberalism." *Loyola Law Review* 61): 69–91.
- Stell v. Savannah-Chatham Board of Education** (1963) 220 F.Supp 667; (1964) 333 F.2d 55.
- Thomas, William** (1982) "Black intellectuals' critique of early mental testing: a little-known saga of the 1920s." *American Journal of Education* 90 (3): 258–92.
- (1984) "Black intellectuals, intelligence testing in the 1930s, and the sociology of knowledge." *Teachers College Record* 85): 477–501.
- Thorndike, Robert L.** (2005) *Measurement and Evaluation in Psychology and Education*, 7th ed. Upper Saddle River, N.J.: Pearson Education.
- Tucker, William H.** (1994) *The Science and Politics of Racial Research*. Urbana: University of Illinois Press.
- (2002) *The Funding of Scientific Racism: Wickliffe Draper and the Pioneer Fund*. Urbana: University of Illinois Press.
- Tyack, David** (1974) *The One Best System: A History of American Urban Education*. Cambridge, MA: Harvard University.
- van den Haag, Ernest** (1960) "Social science testimony in the desegregation cases – a reply to professor Kenneth Clark." *Villanova Law Review* 6 (1): 69–79.
- Weschler, Herbert** (1959) "Toward neutral principles of constitutional law." *Harvard Law Review* 73 (1): 1–35.
- Williams, Robert L.** (1974) "A history of the Association of Black Psychologists: early formation and development." *Journal of Black Psychology* (1): 9–24.
- Yerkes, Robert M., and Helen M. Anderson** (1915) "The importance of social status as indicated by the results of the Point-Scale Method of measuring mental capacity." *Journal of Educational Psychology* 6 (3): 137–50.
- Yudof, Mark G.** (1978) "School desegregation: legal realism, reasoned elaboration, and social science research in the Supreme Court." *Law and Contemporary Problems* 42 (4): 57–110.

**Keith McNamara** has a PhD in history and educational policy from the University of Wisconsin-Madison. His research lies at the intersection of the history of education, legal history, and the history of the social sciences. His dissertation traces the psychological, political, and legal debates that have surrounded testing bias, specifically the idea of "cultural bias" in psychometric testing, from early critiques in the 1920s through the emergence of federal legislation and court decisions in the late 1960s through the 1980s.