The National Coalition on School Diversity

RESEARCH BRIEF 16 | JAN 2023

ACCOUNTABILITY SYSTEMS AND THE PERSISTENCE OF SCHOOL SEGREGATION: RESEARCH EVIDENCE AND FUTURE DIRECTIONS

JAMES NOONAN AND PETER PIAZZA

In recent history, the landscapes of education policy and practice have become increasingly driven by state systems of accountability.[1] Passed in 2001, the federal No Child Left Behind Act (NCLB) codified the use of educational accountability systems that rely on largely quantitative measures – standardized tests, especially – to assess student learning outcomes, rank order districts and schools according to these outcomes, and prescribe sanctions for schools and districts not meeting achievement targets. Meanwhile, American public schools have rapidly resegregated (Frankenberg et al., 2019) since reaching the height of statistical desegregation in 1988 (Orfield et al., 2014).[2] A 2022 Government Accountability Office (GAO) report found that more than one-third of all K-12 public school students attend a school where at least 75 percent of the enrollment comes from the same racial background.

In this brief, we look to recent empirical research to consider *what, if any, relationship exists between systems of accountability and the persistence of school segregation.* Specifically, in the pages that follow, we attempt to answer the following questions:

- *What does empirical evidence suggest about the relationship between contemporary accountability systems and trends in school segregation?*
- *How, if at all, might accountability systems be refined in order to contribute constructively to real integration?*

The shortcomings of standardized testing have long been part of the public debate about education policy. Recently, though, more scholars and advocacy organizations have begun pushing for changes that take aim at persistent racial inequity.[3] For example, the National Education Policy Center released a framework to "center equity and justice and revitalize the original civil rights focus" of the Elementary and Secondary Education Act (ESEA), first passed in 1965 (DeBray et al., 2022).[4] Our brief joins that effort, offering research-based guidance for reflecting on the educational accountability movement and reorienting accountability policy towards integration.

Key Terms and Assumptions

Before we begin, we should clarify some critical terms. First, when we talk about *segregation*, we refer to schools that serve predominantly students or color and/or low-income students as well as schools that serve predominantly White students and/or high-income students. White students have long been the most isolated group in America's public schools and the GAO's recent report confirms this trend. During the 2020-21 school year, nearly half (45%) of all White students attended K-12 schools

1 For a summary of accountability systems in each state, as of December 2021, visit https://www.ecs.org/50-state-comparison-states-school-accountability-systems.

2 We use the term "statistical integration" to refer specifically to school-level racial enrollment. As described further below, we believe that real integration encompasses more than simple school-level demographics.

3 Multiple groups are engaged in efforts to define principles and indicators to inform future ESEA reauthorization. See, e.g. the Opportunity to Learn Commission (https://schottfoundation.org/our-work/otl-policy-commission), National Urban League and UnidosUS (https://nul.org/event/broadening-perspectives-education-assessment-accountability-and-equity), the National Education Association (https://www.nea.org/sites/default/files/2021-02/27604%20GPS%20Indicators%20Framework_Final.pdf), etc.

4 ESEA has been reauthorized numerous times by Congress, sometimes under different names. Both NCLB and the Every Student Succeeds Act (ESSA) are reauthorizations of the 1965 law.

that are at least 75% White. Any discussion of segregation that does not include White and/or high-income segregated schools is therefore missing a critical element of contemporary school segregation.

Second, our use of the term *"real integration"* comes from the 5Rs of Real Integration framework, developed and promoted by youth organizers from IntegrateNYC. Bearing resemblance to the *Green* factors,[5] the 5R framework looks beyond school-level desegregation ("race and enrollment") towards a vision for holistic integration that also includes resource equity, relationships between students and teachers, restorative justice, and representation of diversity among educators.[6] Although we credit the 5R model in our use of the term "real integration," we acknowledge that there are several promising frameworks for measuring real – or holistic – integration and discuss them in more detail later.

Finally, we want to distinguish between *assessment* and *accountability*. Assessment refers to methods of determining what students know and still need to know. Assessment is undertaken through a range of modalities ranging from standardized tests to performance-based assessments and can happen at the classroom-, school- or district-level. Accountability, on the other hand, combines achievement standards with various forms of assessment to make system-level determinations about school performance, and thus quality. Schools that receive high ratings in their state accountability system may be rewarded while those with low ratings may be subject to interventions and sanctions.

It is similarly important to clarify an assumption that guides our review of research and policy initiatives: we believe that the relationship between accountability and segregation is complex, defying simple explanations and solutions. In the debate leading up to the No Child Left Behind Act, civil rights groups were influential in securing provisions for the racial disaggregation of student test score data. Thanks in part to these accountability systems, gaps in achievement by students' race, class, and special education status have been made more transparent.[7] As a result, the range of stakeholders in education systems and in society more broadly have greater insight into the relationship between student test score outcomes and various elements of racially unequal schooling, such as inequity in school funding or access to experienced educators (Cardichon et al., 2020; EdBuild, 2019).

> *"[T]he relationship between accountability and segregation is complex, defying simple explanations and solutions."*

Though racially disaggregated test score results have helped to highlight inequity, underlying systemic inequity remains unchanged.[8] In fact, evidence from the past

5 The "*Green* factors" were first adopted by the federal courts in the 1960s for defining a fully integrated school system. They included measures of equity in facilities; student, faculty and staff assignments; transportation; and extracurricular activities. See *Green v. County School Board of New Kent County*, 390 U.S. 936 (1968).

6 The 5R framework is outlined at https://integratenyc.org/platform.

7 We further acknowledge that the availability of disaggregated data has enabled numerous research studies.

8 Indeed, as others have noted, standardized testing may be an imperfect vehicle for racial equity due, in part, to its origin in the eugenics movement (Muñiz & Barragán, 2022). See also: https://www.nea.org/advocating-for-change/new-from-nea/racist-beginnings-standardized-testing.

two decades of state accountability clearly illustrate the shortcomings of the public release of data as a policy intervention (Hutt & Polikoff, 2020; Rogers, 2006).[9] Put simply, the widespread availability of racially disaggregated accountability data has not led to the large-scale improvements in academic outcomes that motivated civil rights support – including support from the school integration community – for an NCLB-style testing regime (see Reid, 2005).[10]

Relatedly, as detailed in the following section, the most common forms of accountability have created new barriers to racial inclusion in American public education. Narrow and flawed measures have taken on outsized significance in how we measure the quality of our schools and even our neighborhoods.[11] Especially considering the way that standardized test scores tend to mirror school demographics,[12] terms like "good schools" and "bad schools" have not only become commonplace in the educational debate, they have also functioned as proxies for school racial composition (Knoester & Au, 2017; Muñiz & Barragán, 2022; Noonan & Schneider, 2022; Piazza, 2022).

"Especially considering that standardized test scores tend to mirror school demographics, terms like 'good schools' and 'bad schools' have not only become commonplace in the educational debate, they have also functioned as proxies for school racial composition."

The Relationship Between Accountability and Segregation

To understand the relationship between accountability systems and school segregation, we first consider the effects of desegregation on student learning and how – if at all – those effects changed with the introduction of high-stakes accountability. Research literature is clear on the first question: desegregation efforts were highly effective at improving student learning and narrowing the test score gap between Black and White students (Johnson, 2019; reardon et al., 2015). Regarding the second question, the advent of high-stakes accountability appears to be associated with accelerated resegregation and the stalled narrowing of racial achievement gaps.

Desegregation efforts predating high-stakes testing and accountability appeared to drive meaningful improvements in student

9 For a more in-depth discussion of the shortcomings of "mandated disclosure" across domains, see Ben-Shahar and Schneider (2010).

10 For more on the legacy of the test-based accountability movement, see Ravitch et al. (2022).

11 Reliance on test-based measurement has also led to flawed language that can act as an additional barrier to racial justice. For example, recent research has found that when news coverage uses the term "achievement gap," viewers embrace stronger anti-Black narratives (Quinn, 2020).

12 Standardized test scores are strongly influenced by out-of-school variables at the family and community level (e.g. family income, poverty/wealth concentration, etc.), thus they often reflect the racial and socioeconomic makeup of a school (Haertel, 2013; Tienken et al., 2017). For this reason, Schneider (2017) has referred to school ratings heavily based on standardized testing as "demographic data in disguise" (see also Koretz, 2017).

learning.[13] Using long-term trends from the National Assessment of Educational Progress (NAEP), a no-stakes standardized test given regularly to a nationally representative sample of US students, Grissmer and colleagues (1998) noted rapid improvements among Black students during the 1970s and 1980s and attributed these changes, in part, to desegregation and civil rights initiatives that improved schooling conditions for Black students.[14] Hanushek and Raymond (2005) similarly found that pre-NCLB accountability systems had positive effects on Black student learning outcomes but had greater positive effects on White students, thus widening the Black-White test score gap. Many of these accountability systems were developed and introduced by states in the 1990s, when increasing numbers of districts were released from court desegregation orders (reardon et al., 2012). Meanwhile, Hanushek and Raymond (2005) found that racially isolated schools had larger negative effects on the learning of Black students than on White or Latinx students, suggesting that the resegregation of schools was disproportionately affecting Black student learning.

The passage of the No Child Left Behind Act in 2001 reinforced – and may even have accelerated – trends toward resegregation, without appreciably improving learning outcomes for students. Ryan (2004) noted the "incentives to segregate" (p. 961) embedded in the law. Specifically, schools now had an incentive to exclude students who tended to score lower on standardized tests, including students of color and students experiencing poverty (Knoester & Au, 2015). Indeed, some empirical evidence suggests that public schools operating under high-stakes accountability systems have taken active or tacit steps to boost test scores by managing their student populations: for example, excluding students from testing who were more likely to score lower (Vasquez Heilig & Darling-Hammond, 2008), drawing attendance zone boundaries to exclude students of color and/or low-income students (Wells & Holme, 2005), or "pushing out" students based on disciplinary records (Kho et al., 2022). Importantly, the incentive to keep low-scoring students out of "high-performing" schools helped render one of the most integration-focused aspects of the law – the school transfer provision[15] – largely meaningless.

> *Under NCLB, "[S]chools now had an incentive to exclude students who tended to score lower on standardized tests, including students of color and students experiencing poverty."*

In addition, state accountability systems often reward schools based on student passing rates in advanced courses; yet, recent research has shown that students in racially segregated schools are less likely

13 Generally speaking, "high-stakes" tests have consequences attached to performance, such as the consequences written into NCLB for schools or districts that failed to make "adequate yearly progress" on state standardized tests. Conversely, "low-stakes" or "no-stakes" tests assess performance at a single point or track long-term trends but do not mandate policy changes.

14 Recently-released NAEP data indicate that racial test-score gaps have widened in recent years, likely due to existing inequities that were further exacerbated during the pandemic. For more, visit https://nces.ed.gov/nationsreportcard.

15 Under NCLB, Title I schools that did not make adequate yearly progress for two or more consecutive years were required to allow their students to transfer to a higher-rated school in the district. See https://www2.ed.gov/rschstat/eval/choice/nclb-choice-ses-final/index.html.

to have access to Advanced Placement (Government Accountability Office, 2016) as well as Gifted and Talented programs (School Diversity Advisory Group, 2019). Meanwhile, research also indicates that schools with advanced coursework may use gatekeeping mechanisms to create within-school segregation under the banner of so-called ability grouping (Lewis & Diamond, 2015).

Additionally, scoring and ranking schools according to students' performance on tests appears to have shaped perceptions of "good" and "bad" schools, which in turn may affect families' choices about where to live and where to enroll children in school (Miner, 2020; Schneider et al., 2018; Schneider et al., 2020). Meanwhile, NCLB's short-term positive effects on student learning were generally modest (Dee & Jacob, 2011; Figlio & Loeb, 2011; Lee, 2008).

Looking beyond school policy specifically, educational accountability systems appear to exacerbate housing segregation, which has a direct impact on the racial and economic resegregation of schools. In particular, the use of publicly prominent "school report cards," which pre-date NCLB (Portz & Beauchamp, 2020), may shape public perceptions about school quality. Figlio and Lucas (2004) found that an "A" on a school report card had significant effects on housing prices and nearby property values (see also Hasan & Kumar, 2019). Consistent with these findings, reardon and colleagues (2019) determined that geographic variation in racial achievement gaps were largely explained by differences in family income and school segregation.

> *"Figlio and Lucas (2004) found that an 'A' on a school report card had significant effects on housing prices and nearby property values."*

The acceleration of post-NCLB resegregation was especially pronounced in states that had NCLB-like accountability systems prior to the law's passage (Davis et al., 2015).[16] Wells and Holme (2005) collected contemporary data on six schools that had been racially mixed in the 1970s, and found that – 20 years later – five of the six schools had either lost all or nearly all of their White students or were in the process of resegregating. "In all five of these sites," they wrote, "state-mandated tests appear to have played a role in this White and middle- and upper-middle- class flight" (p. 187). Each of these resegregating levers operate within a broader sociopolitical context in which implicit racial bias, especially anti-Blackness, persists in educational settings (Chin et al., 2020; Pearman, 2021). Implicit bias also appears to influence decisions – often by White and/or affluent families – about where to live and where to send their children to school (Billingham & Hunt, 2016; Freidus, 2022; Holme, 2002).

The resegregation of schools is an obstacle to student learning. Early in the NCLB era, Borman et al. (2004) found that accountability-era policy efforts to close racial achievement gaps in Florida – including reformed educational funding and punitive measures against schools that failed to meet performance benchmarks – had been largely unsuccessful, leading the authors to urge "great caution in abandoning school integration as a mechanism to close the

16 In 2002, when NCLB was enacted, 29 states had educational accountability systems that tied consequences to school performance. However, there were wide variations in the nature of testing, the extent that test scores were disaggregated by student subgroups, and the range of measures used beyond test scores. For more, see https://aheadoftheheard.org/accountability-before-during-and-after-nclb.

racial gap" (p. 626). More recently, reardon and colleagues (2022), using longitudinal data from state accountability tests in all 50 states, continued to find a strong association between racial segregation and the persistence of racial achievement gaps, largely due to disproportionate exposure to concentrated poverty.

The impact of educational accountability on student learning outcomes should be viewed in the context of the inequitable distribution of resources afforded to schools (EdBuild, 2019), which long predates both the movement for integration and the movement for accountability. Compared to segregated schools serving students of color, racially isolated White schools tend to have far more of the resources necessary for effective teaching and learning – resources such as experienced and effective teachers (Benson et al., 2020; Cardichon et al., 2020), access to appropriate mental and emotional support personnel (Locke et al., 2017), access to culturally responsive curriculum (NYC Coalition for Educational Justice, 2019), and school district expenditures (Sosina & Weathers, 2019).

In addition, schools targeted for punitive consequences under accountability systems – often those serving low-income students and/or students of color – have tended to respond with academically unproductive interventions like narrowed curriculum and increased test preparation (e.g., Booher-Jennings, 2005; Diamond & Spillane, 2004; Koretz, 2008). Rectifying resource and opportunity gaps between racially isolated schools were, of course, a primary motivator for the desegregation movements in the mid-20th century. And yet, after more than 20 years of educational accountability at the federal level, racial achievement and opportunity gaps have persisted and even widened (reardon et al., 2022).

"The impact of educational accountability on student learning outcomes should be viewed in the context of the inequitable distribution of resources afforded to schools, which long predates both the movement for integration and the movement for accountability."

Indeed, the ethos and impacts of accountability have surprising commonalities with some of the after-effects of *Brown v. Board of Education*. For example, each has been characterized by the firing or voluntary attrition of teachers, the closing or restructuring of schools, and a narrowing of the curriculum (Walker, 2018). The question then becomes: What, if anything, can be done to disrupt this relationship?

THE FUTURE OF ACCOUNTABILITY AND REAL INTEGRATION

As evident in the research literature, test-based accountability systems have become embedded in the social structures that shape students' access to schools and neighborhoods. In this section, we point towards a variety of potential tools – both within and parallel to current federal K-12 policies – that might chip away at this relationship. We first discuss the limited potential for real integration in existing federal law, then review efforts to rethink accountability underway at the state and local levels. We conclude by highlighting strategies and existing resources for reorienting accountability towards school integration. While it is beyond the scope of this brief to describe any individual solution in detail, we hope this list can be a helpful resource for conversations about how we might move towards a future of accountability that facilitates racial equity and integration.

Existing Levers at the Federal Level

The Every Student Succeeds Act (ESSA), which replaced NCLB in 2015, largely kept NCLB's testing regime in place while giving states more room to define their own measures of success, including approaching school accountability beyond test scores.[17] Groups like the Mexican American Legal Defense and Educational Fund (2007), the NAACP Legal Defense Fund (2015), and others in the civil rights community, applauded the passage of NCLB and, later, ESSA for helping to provide "critical information about existing and pervasive disparities [that] helps to identify areas to target interventions."[18] ESSA also required states to define and measure school quality using at least one **non-academic indicator of student success**, known as the "**5th indicator**."[19] In 2016, the National Coalition on School Diversity (NCSD) suggested that states include indicators that measure progress on integration in their ESSA accountability systems (Hilton, 2016).

17 Although many ESSA provisions could be analyzed for their potential to promote school integration (See, for example: http://www.school-diversity.org/wp-content/uploads/NCSDPB10_Final.pdf), this brief focuses specifically on ESSA provisions related to school quality measurement for the purposes of state accountability.

18 LDF's 2015 Statement on the Every Student Succeeds Act is available at https://www.naacpldf.org/press-release/ldf-statement-on-the-every-student-succeeds-act.

19 ESSA § 1111(c)(4)(B)(v)(I). As stipulated in ESSA, the non-academic indicator must a) allow for "meaningful differentiation in school performance" and b) must be based on measures that are "valid, reliable, comparable and statewide." Others have referred to this as the "opportunity indicator." See, for example: https://www.nea.org/resource-library/essa-improves-state-designed-accountability-systems.

However, few state accountability systems have truly embraced the 5th indicator as an opportunity to pursue racial equity, much less as a vehicle for real integration.[20] Moreover, any 5th indicator is hampered by ESSA's requirement that academic measures be assigned "much greater weight" in accountability determinations than non-academic measures. Research from Portz and Beauchamp (2020) found that, in some states, weighting for the 5th indicator is so low that it effectively has no impact on the state accountability system.[21] Unsurprisingly, then, states' implementation of the law has not led to meaningful accountability changes necessary to help drive equity (National Urban League, 2019).

Nonetheless, were states so inclined, the 5th indicator provides some room for them to incorporate indicators that are less influenced by out-of-school factors and are more capable of driving systemic change. For example, a comparatively smaller number of states include "equity" and/or "school climate" indicators in their ESSA plans. Recent research suggests that these kinds of measures could surface information about a school's strengths and areas for growth that are less tied to out-of-school factors than standardized test scores. Schneider and colleagues (2021b) looked at how school accountability outcomes in Massachusetts would be different if the state used student school climate survey data as 25% of its state accountability system. Their findings indicated that schools serving predominantly Black and Latinx student populations would see their rankings rise – sometimes by as much as 10 percentile points – when using student perception surveys.

In 2018, NCSD reported that only one state proposed using measures of real integration for its 5th indicator. New York's ESSA plan identifies "integration of students" as one of several potential indicators in an accountability system that utilizes multiple measures of school quality. Under this plan, which was formally approved in 2018, the state would consider the extent to which students of various racial and socioeconomic subgroups "are in schools and classrooms together," compared to their presence in the district as a whole. Importantly, this information would factor into a school's overall accountability rating.[22] To date, however, officials have not made use of this indicator in their approach to accountability.

Even if more states were interested in using the 5th indicator as a vehicle for real integration, there is little guidance about

20 The majority of states use "chronic absenteeism" as their non-academic indicator and several use measures of student behavior, school discipline, or so-called "dropout" rates, which – just like standardized test scores – are strongly influenced by out-of-school variables and can be susceptible to bias (Education Commission of the States, 2018; Portz & Beauchamp, 2020). Further, the U.S. Department of Education has encouraged states to use Opportunity to Learn indicators as part of their academic and non-academic measures of school performance. See USED guidance letter here: https://oese.ed.gov/files/2022/09/Assessment-Letter-FINAL_Redacted-9-2022.pdf. Holistic school quality indicators – including Opportunity to Learn measures – can also be used as part of state-required school improvement plans.

21 A report from the Education Commission of the States found that accountability systems often weigh the 5th indicator as about 10-15% of the school's accountability ranking. Some states use the 5th indicator as a "tie-breaker" between schools with otherwise similar ratings and some even use it to incentivize participation in standardized tests (e.g., Michigan). See https://reports.ecs.org/comparisons/states-school-accountability-systems-2021.

22 To view New York's full ESSA Plan, visit: http://www.nysed.gov/essa/nys-essa-plan.

specific measures that might be used. Especially in the absence of federal direction, the civil rights and research communities could support racial equity in school quality measurement and accountability systems by developing robust indicators aligned with the 5Rs – or a related framework – and by providing states and districts with support on their implementation. As described further below, NCSD has begun developing a tool for measuring real integration in interdistrict school integration programs. NCSD (2020) and its partner organizations have also developed a model state policy that would require states to include measures of racial and socioeconomic integration in annual school ratings. The model policy includes clear guidance on how to measure integration via a proportionality score and how to ensure that historically underserved student subgroups have equitable access to school supports associated with high educational achievement.

"[T]he civil rights and research communities could support racial equity in school quality measurement and accountability systems by developing robust indicators aligned with the 5Rs – or a related framework – and by providing states and districts with support on their implementation."

In addition to its non-academic indicator, ESSA includes grant funds and flexibility provisions similarly aimed at moving beyond standardized tests in school quality measurement. The **Competitive Grants for State Assessments (CGSA)**[23] program provides funding for states to use multiple measures in assessing student learning, such as performance-based assessment. States might also experiment with using academic growth scores, which – unlike proficiency rates – are less reflective of a school's demographic composition (Angrist et al., 2022; Houston & Henig, 2021).[24] As above, however, the law also includes restrictions that dampen the potential for experimentation in the CGSA program. Although this provision would allow states to use more culturally-responsive forms of student assessment, the use of alternative assessments would not alter the calculus of school accountability rankings. Importantly, the CGSA funding may also be a barrier to large-scale implementation of a high quality performance assessment program. For example, New York state spent $10.3 million on standardized testing in 2012, yet the state only received $3 million in CGSA funding (Chingos, 2012). Similar to the 5th indicator, constraints built into ESSA may limit the extent to which the CGSA program can be scaled up to the state level in a way that captures the nuance of student learning and advances racial equity.

Finally, the **Innovative Assessment Demonstration Authority (IADA)** component of ESSA gives states flexibility from the requirement that all students in a state are evaluated using the same form of assessment. By implementing pilot efforts at a smaller scale, IADA gives states freedom to experiment, but again within certain constraints: scores from IADA

23 For more information on the Competitive Grants for State Assessments program, visit: https://oese.ed.gov/offices/office-of-formula-grants/school-support-and-accountability/competitive-grants-for-state-assessments.

24 We acknowledge that using growth scores comes with its own challenges, including very practical time and resource constraints that could negatively impact students, families, and educators. The pros and cons of new approaches should be carefully deliberated by stakeholders.

efforts must be comparable to traditional assessments for students in the aggregate as well as student subgroups. ESSA allows up to seven states to take advantage of this flexibility. As of January 2023, five states had been approved.[25] Interviews with IADA leaders in the implementing states indicate that the promise of the pilot efforts are limited by the constraints of the law, especially requirements that new forms of assessment be eventually scaled up to statewide implementation (Marion & Evans, 2021). Recently, in fact, the state Superintendent of Georgia announced that IADA efforts are stalled in his state. He described IADA regulations "a very difficult, uphill climb," referring specifically to the comparability requirement[26] in ESSA (Amy, 2022).

In proposing a civil rights framework for a future reauthorization of ESEA, DeBray and colleagues (2022) wrote that the law "could and should provide strong supports for integration and racial equity programs" (p. 17). As we describe throughout this brief, assessment policy is a critical part of this effort. Along those lines, a future iteration of the law could incentivize equity-oriented changes to state assessments by relaxing its most onerous requirements. Referring to the school transfer provisions in NCLB, Dimyan-Ehrenfeld (2009) pointed out that "Congress has the opportunity to be creative in the restructuring of these provisions" (p. 246).

We believe the same is true for the non-academic indicator as well as the CGSA and IADA provisions of ESSA. Further research directions, which we discuss below, may provide empirical support for such revisions.

> *"Especially given the ways that state accountability and related sanctions can maintain or exacerbate segregation, it is important to look beyond federal law to understand how states or local coalitions might use alternative forms of accountability to weaken the relationship between school measurement and school composition."*

Parallel Efforts at the State or Local Level

Any flexibility within ESSA operates underneath the law's most central elements: test-based school accountability, ranking, and high-stakes consequences. Especially given the ways that state accountability and related sanctions can maintain or exacerbate segregation, it is important to look beyond federal law to understand how states or local coalitions might use alternative forms of accountability to weaken the relationship between school measurement and school composition. Professional organizations have offered visions for a broader approach to assessment and accountability, while district and non-profit advocates have begun

25 States approved for IADA flexibility: Louisiana & New Hampshire (in 2018), Georgia & North Carolina (in 2019) and Massachusetts (in 2020). See https://www2.ed.gov/admins/lead/account/iada/index.html and https://www.edweek.org/policy-politics/massachusetts-gets-green-light-to-pilot-innovative-science-assessment/2020/04.

26 The U.S. Department of Education's IADA flexibility application requires that "The innovative assessment system demonstrates comparability to the statewide assessments." Put differently, proficiency determinations based on IADA pilot assessments should not differ dramatically from annual, student-level proficiency determinations based on the traditional form of standardized assessment (see Marion & Evans, 2021). Access the IADA application at https://www2.ed.gov/admins/lead/account/iada/essainnovativeassessapp121718.pdf.

similar experiments at the local level.[27] Instead of focusing on narrow measures of academic learning, state and/or district efforts can be more holistic, and their low-stakes nature allows these efforts the freedom to experiment. Eventually, low-stakes implementation can drive larger improvements in how state accountability systems make use of promising alternative assessment strategies such as holistic school quality measurement and the use of performance assessments in measuring student learning (Darling-Hammond & Falk, 2013; Famularo et al., 2018).

Unlike federal-level accountability, local forms of school accountability more commonly include measures of school inputs alongside consideration of student learning outcomes. As a result, these efforts can be more attuned to metrics that are particularly relevant to the contemporary movement for school integration, such as opportunities for culturally-responsive education or a student's sense of belonging in their school community. Alternative forms of accountability can also be powerful modes for understanding school quality in that they often include an essential (though woefully overlooked) source of school quality data: the perspectives of students and teachers who animate school communities.[28] Typically these voices are avoided in educational policy discourse in favor of so-called "objective" test-based measurement; however, as we have shown, this approach has not delivered on promises for equity.

Though the movement for accountability beyond test scores is still in its early stages, several initiatives offer insight into how such parallel efforts might be undertaken and sustained. A recent report from the Beyond Test Scores project at UMass Lowell detailed promising practices from various state and local accountability efforts (Carey & Schneider, 2022). For example, the Massachusetts Consortium for Innovative Education Assessment (MCIEA) measures school quality according to 34 indicators, including students' perspectives on cultural responsiveness in curriculum. MCIEA's full data collection portfolio – including its survey scales and administrative data measures – are publicly available as source material for other states, districts, or coalitions pursuing similarly holistic forms of school quality measurement. In addition, as with state accountability systems, MCIEA data are posted publicly on an online dashboard.[29]

> *"Across the country, there appears to be a growing appetite for alternative approaches to school quality measurement."*

Across the country, there appears to be a growing appetite for alternative approaches to school quality measurement. Results from large scale surveys designed to gather public feedback about school accountability in Texas and North Carolina attest to the shift in public attitudes. Of more than 26,000

27 *See, e.g.* the National Education Association's Principles for the Future of Assessment, the Position Statement on High Stakes Testing by the American Educational Research Association, and a Position on High-Stakes Assessment by the National Association of Secondary School Principals.

28 See Schneider et al. (2021a) for a toolkit on developing a holistic school quality measurement system, including a catalog of free, open-source survey instruments. The following resources also offer models for incorporating community voice in school quality measurement: https://www.nea.org/student-success/great-public-schools/community-schools/what-are-they; https://lovingcities.schottfoundation.org.

29 To explore the MCIEA School Quality Measures Dashboard, visit https://mciea-dashboard.herokuapp.com/welcome.

survey respondents in North Carolina,[30] 90% indicated that they favor an accountability system that "includes measures beyond test scores" (Sorrells, 2022). Meanwhile, in a survey of more than 15,000 Texas residents,[31] just 1% of respondents said that they believe the state test is a very accurate measure of school quality (Raise Your Hand Texas, 2022).

In addition to providing a broader range of data to drive real integration in schools and districts, alternative forms of accountability also contribute to continuing research on the benefits of school integration. A fuller range of school quality data can generate more evidence about student experiences in schools and challenge the "good" schools versus "bad" schools binary reinforced by test-based measurement. For example, using MCIEA survey responses from over 25,000 students, Schneider and colleagues (2020) compared student experience in "diverse"[32] schools with those in comparatively more segregated schools. They found that White students in diverse schools reported more positive experiences than their White counterparts in more homogeneous schools, including higher levels of physical safety, engagement, sense of belonging, and civic participation. Consistent with the recommendations from major school integration advocacy organizations (Potter et al., 2021), research in this vein helps illustrate the benefits of school integration for all students, and it would not be possible without alternative forms of school quality measurement.

"Schneider and colleagues (2002) ....found that White students in diverse schools reported more positive experiences than their White counterparts in more homogeneous schools, including higher levels of physical safety, engagement, sense of belonging, and civic participation."

While there is reason to be hopeful about local level accountability efforts, there is still much room for growth in connecting alternative accountability measures more explicitly to the values that drive the contemporary movement for school integration. As described above, the 5R framework offers a comprehensive vision for school integration. Just like alternative forms of accountability aim for a more holistic form of school quality measurement, the 5Rs – and related models – aim for a more holistic approach to school integration. It seems natural, then, that these efforts be combined in the form of a school quality measurement framework oriented towards real integration. NCSD embarked on an effort along these lines during the summer of 2022. Specifically, the coalition began building a framework for school quality measurement that is designed to support students of color participating in interdistrict integration programs. These programs were created to address segregation across metro areas, thus participating schools often serve a predominately White student population. As above, this kind of measurement adds complexity to our understanding of schools that are often simply labeled as "good" based on their test scores.

30 To see the full survey results from North Carolina, visit https://publicinput.com/Report/kym1l4l4axl.

31 See full survey results from Texas at https://www.raiseyourhandtexas.org/advocacy/futureoftexastesting.

32 Schneider and colleagues (2020) define a school as "diverse" if it serves no more than 75% of students from one racial subgroup and at least 25% White students. For more on this definition, visit https://cecr.ed.psu.edu/sites/default/files/Demography_Report_FINAL_7.24.20.pdf.

As we consider what a school accountability model that facilitates real integration should look like, several other resources are worth mentioning:

- Prior to closing its doors, the Reimagining Integration: Diverse and Equitable Schools (RIDES) project worked with Panorama Education to launch an "Equity & Inclusion" survey[33] aimed at measuring progress on "commitment to dismantling racism and oppression" and "diversity," two aspects of RIDES's ABCDs of Real Integration[34] framework.

- In Despite the Best Intentions: How Racial Inequality Thrives in Good Schools, Lewis and Diamond (2015) highlight strategies for understanding within-school segregation, especially in majority White schools.

- A recent doctoral dissertation by Olivia Hussey, Measuring What Matters: Developing an Assessment Tool to Measure 'Real Integration' in Schools (2020), similarly builds out a model for measuring real integration based on the 5Rs of Real Integration framework.

- The Future Ready: Integrated Schools Continuum of Excellence, developed by Carole Learned-Miller, Kimberly Bridges, and Reginald Johnson (2016), is a rubric designed to help schools and systems identify areas for growth in their efforts to reduce racial isolation, prejudice, and opportunity gaps and build 21st century skills.[35]

33 To view Panorama Education's "Equity and Inclusion" survey and survey implementation materials, visit: https://www.panoramaed.com/equity-inclusion-survey.

34 The ABCD framework consists of: academics, belonging, commitment to dismantling racism and oppression, and diversity. To learn more about the ABCD integration framework, visit: https://rides.gse.harvard.edu/abcds.

35 This continuum was originally published on the Transforming Teaching website. Learn more at https://www.totransformteaching.org/integrated-schools-prepare-future-ready-students. Access the continuum at https://drive.google.com/file/d/1IhE5ssZbI-Ux-gv9I55VY-S34oI5CT0W/view.

CONCLUSION AND RECOMMENDATIONS

The late Richard Elmore (2003) argued that accountability systems, far from being coherent and evidence-based, were instead "political artifacts crafted…to meet the demands of political action" (p. 295). In the case of the US, which soured early in its commitment to real integration, educational accountability systems were developed with the tacit assumption among policymakers that "we can equalize schools without dealing with segregation" (Orfield & Lee, 2007, p. 7). Based on our review of the research evidence, we conclude that testing and accountability – in the absence of complementary and sustained integration efforts – have not lived up to that promise.

On the contrary, we find that testing and accountability systems may well be doing more to obstruct than ease the path toward equitable education. Desegregation efforts in the 1960s and 1970s were positively associated with an improvement in schooling conditions and learning outcomes for all students. Accountability systems like those mandated by NCLB and ESSA, on the other hand, have had a more mixed record, with incremental gains for some students, in some subjects, in some grades. Early in the standards and accountability era, some argued that public reporting of racially disaggregated test scores would lead to greater equity in school funding and academic resources by identifying and targeting those schools that needed the most support. However, the best intentions underlying this promise are still a long way from being realized in America's classrooms. Indeed, in the last several decades, resource inequities across schools and school systems have only grown more stark.

We further find that the levers available to promote the academic and social benefits of real integration within these accountability systems are limited. Federal education law and regulations allow for some flexibility in measuring school quality, but they maintain an emphasis on proficiency scores, which are strongly influenced by out-of-school variables. Thus, they often reflect the racial and socioeconomic makeup of a school. Further, provisions oriented towards moving beyond test scores are constrained by onerous regulations that give preferences to statewide comparability and school ranking. Meanwhile, local level initiatives to assess school quality more holistically are underway in several states and could, over time, serve as existence proofs for the concept that states and districts can incorporate measures of real integration into accountability systems. Even if parallel systems don't reach the level of federal or state policy, they provide schools with a broader spectrum of meaningful and relevant data to create more inclusive learning environments, and these data can generate new research that complicates narrow and overly simplistic conceptions of school quality.

We urge that education policy related to accountability – and the extent to which accountability might promote or hinder real integration – be informed by the best research evidence available. To strengthen the research base on school accountability and real integration, we offer the following recommendations:

First, because outcomes of accountability systems are in part a reflection of the measurement tools they employ, **we urge further research on the development and potential impact of new school quality frameworks and measures.** For example, there is a considerable amount of unexplored potential in the development of culturally-relevant performance assessments as well as the use of growth scores and/or student-centered survey data in state accountability systems. Alternative forms of school quality measurement, however, are relatively new and more information is needed to better understand how accountability systems would (or would not) change with the incorporation of new measures.

Second, as grassroots initiatives to broaden school quality measurement and accountability take shape, **research is needed to assess the impact of these initiatives and how – if at all – they might complement or even replace local, state, or federal accountability systems.** We detail evaluation and empirical research on MCIEA above, but similar efforts are underway with consortia in California, Colorado, New York, and more (see Carey & Schneider, 2022). These are sites for experimentation in school accountability as well as settings for research on alternative forms of accountability. With access to a wider array of school quality data, researchers, for example, can ask questions about how parents use new information in their school choice decisions or about how school or district leaders are responding to survey data indicating that students of color do not feel fully included in the school community. Ultimately, more holistic data emerging from these initiatives can help shape public perceptions of "good" and “bad” schools in a way that contributes constructively to the movement for real integration.

Finally, too often research efforts on accountability and on segregation are siloed, but as this research brief helps make clear, these two domains are in fact closely interwoven. While research exists on the relationship between segregation and accountability broadly speaking (e.g., reardon et al., 2022), there are fewer studies exploring the way that accountability systems interact with the many residual impacts of segregation such as funding inequities, teacher shortages, curriculum quality, school closures, or parental decision-making. **Studies that trace the downstream effects of accountability within and across schools and communities, including its effect on residential segregation and school funding disparities, could lead to new conceptual frameworks and language for talking about racial equity.**

WE RECOMMEND RESEARCH ON:

- The development and potential impact of new school quality frameworks and measures
- The impact of grassroots initiatives to broaden school quality measurement and assessment, and how they might complement or even replace local, state, or federal accountability systems
- The downstream effects of accountability within and across schools and communities, including its effect on residential segregation and school funding disparities

This brief was written by James Noonan and Peter Piazza, with support from the National Education Association.

ABOUT THE AUTHORS

James Noonan, Ed.D., is an assistant professor in the McKeown School of Education at Salem State University in Salem, MA. Prior to joining the faculty at SSU, he was the Director of School Quality Measures for the Massachusetts Consortium for Innovative Education Assessment (MCIEA). His research interests include the intersection of teacher identity and teacher learning as well as the relationship between school quality and equity.

Peter Piazza, Ph.D., is the Director of School Quality Measures for the Education Commonwealth Project, an effort to disseminate MCIEA's school quality measurement system to districts across Massachusetts. He is also a member of the National Coalition on School Diversity through the School Diversity Notebook, where he writes about news events and new research in the contemporary struggle for school integration.

GRATITUDE

In 2022, NCSD conducted one-on-one conversations with leaders in the field of educational assessment and K-12 accountability policy. Those conversations shaped our arguments throughout this brief, and we are appreciative of those who made the time to share their perspectives with us. Special thanks, also, to Christine DonFrancesco at the National Education Association for her helpful feedback. Last, we thank Jack Schneider, Erica Frankenberg, Philip Tegeler, Gina Chirichigno, and Jenna Tomasello for their insights and suggestions.

Last updated: January 12, 2023

SOURCES CITED

Amy, J. (2022, October 1). Efforts falter to develop new Georgia K-12 state tests. The Associated Press. https://www.wsav.com/news/efforts-falter-to-develop-new-georgia-k-12-state-tests

Angrist, J., Hull, P., Pathak, P. A., & Walters, C. R. (2022). Race and the Mismeasure of School Quality (No. w29608). National Bureau of Economic Research. https://blueprintlabs.mit.edu/research/race-and-the-mismeasure-of-school-quality

Ben-Shahar, O., & Schneider, C. E. (2010). *The Failure of Mandated Disclosure* (John M. Olin Program in Law and Economics Working Paper No. 516). The Law School, University of Chicago.

Benson, T. A., Bryant, A., & Gezer, T. (2020). Segregation within integrated schools: Racially disproportionate student-teacher assignments in middle school. *Education Policy Analysis Archives*, 28, 170. https://doi.org/10.14507/epaa.28.5503

Billingham, C. M., & Hunt, M. O. (2016). School racial composition and parental choice: New evidence on the preferences of white parents in the United States. *Sociology of Education*, 89(2), 99-117. https://doi.org/10.1177/0038040716635718

Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas Accountability System. *American Educational Research Journal*, 42(2), 231 – 268. https://doi.org/10.3102/00028312042002231

Borman, K. M., Eitle, T. M., Michael, D., Eitle, D. J., Lee, R., Johnson, L., Cobb-Roberts, D., Dorn, S., & Shircliffe, B. (2004). Accountability in a postdesegregation era: The continuing significance of racial segregation in Florida's schools. *American Educational Research Journal*, 41(3), 605 – 631. https://doi.org/10.3102/00028312041003605

Cardichon, J., Darling-Hammond, L., Yang, M., Scott, C., Shields, P. M., & Burns, D. (2020). *Inequitable Opportunity to Learn: Student Access to Certified and Experienced Teachers*. Learning Policy Institute. https://learningpolicyinstitute.org/product/crdc-teacher-access-report

Carey, A., & Schneider, J. (2022). *Alternative Educational Accountability Right Now: Examples from the Field*. Beyond Test Scores Project, UMass Lowell. https://static1.squarespace.com/static/60aea369b2b1517a8d2afb0c/t/628789812218ce48238e6c94/1653049733099/Alt_Edu_Accountability_Final.pdf

Chin, M. J., Quinn, D. M., Dhaliwal, T. K., & Lovison, V. S. (2020). Bias in the air: A nationwide exploration of teachers' implicit racial attitudes, aggregate bias, and student outcomes. *Educational Researcher*, 49(8), 566 – 578. https://doi.org/10.3102/0013189X20937240

Chingos, M. M. (2012, November). Strength in Numbers: State Spending on K-12 Assessment Systems. Brookings Institution. https://www.brookings.edu/wp-content/uploads/2016/06/11_assessment_chingos_final_new.pdf

Darling-Hammond, L., & Falk, B. (2013). Teacher Learning through Assessment: How Student-Performance Assessments Can Support Teacher Learning. Center for American Progress. https://www.americanprogress.org/wp-content/uploads/2013/09/TeacherLearning.pdf

Davis, T., Bhatt, R., & Schwarz, K. (2015). School segregation in the era of accountability. *Social Currents*, 2(3), 239 – 259. https://doi.org/10.1177/2329496515589852

DeBray, E., Finnigan, K.S., George, J., & Scott, J. (2022). A Civil Rights Framework for the Reauthorization of ESEA. National Education Policy Center. http://nepc.colorado.edu/publication/reauthorization

Diamond, J. B., & Spillane, J. P. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality. *Teachers College Record*, 1145 – 1176.

Dimyan-Ehrenfeld, J. (2009). Making lemonade: Restructuring the transfer provisions of the No Child Left Behind Act. *Georgetown Journal on Poverty Law & Policy*, 16(1), 217-248.

SOURCES CITED

Ed Build. (2019, February). $23 Billion. https://edbuild.org/content/23-billion/full-report.pdf

Education Commission of the States. (2018, May). Accountability and reporting: ESSA Plans. http://ecs.force.com/mbdata/mbQuest5E?rep=SA172

Elmore, R. F. (2003). The Problem of Stakes in Performance-Based Accountability Systems. In S. H. Fuhrman & R. F. Elmore (Eds.), *Redesigning Accountability Systems for Education* (pp. 274 – 296). Teachers College Press.

Famularo, J., French, D., Noonan, J., Schneider, J., Sienkiewicz, E. (2018). Beyond Standardized Tests: A New Vision for Assessing Student Learning and School Quality. Center for Collaborative Education. https://www.cce.org/uploads/files/MCIEA-White-Paper_Beyond-Standardized-Tests.pdf

Figlio, D., & Loeb, S. (2011). School Accountability. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbooks in Economics* (Vol. 3, pp. 383 – 421). Elsevier. https://doi.org/10.1016/B978-0-444-53429-3.00008-9

Figlio, D. N., & Lucas, M. E. (2004). What's in a grade? School report cards and the housing market. *The American Economic Review*, 94(3), 591 – 604.

Frankenberg, E., Ee, J., Ayscue, J. B., & Orfield, G. (2019). *Harming our common future: America's segregated schools 65 years after Brown*. The Civil Rights Project/Proyecto Derechos Civiles. https://escholarship.org/content/qt23j1b9nv/qt23j1b9nv.pdf

Freidus, A. (2022). Segregation, diversity, and pathology: School quality and student demographics in gentrifying New York. *Educational Policy*, 36(4), 822 – 848. https://doi.org/10.1177/08959048221087206

Grissmer, D., Flanagan, A., & Williamson, S. (1998). Why Did the Black-White Score Gap Narrow in the 1970s and 1980s? In C. Jencks & M. Phillips (Eds.), *The Black-White Test Score Gap* (pp. 182 – 226). Brookings Institution Press.

Haertel, E. H. (2013). *Reliability and Validity of Inferences about Teachers Based on Student Scores*. Educational Testing Service. https://www.ets.org/Media/Research/pdf/PICANG14.pdf

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24(2), 297 – 327. https://doi.org/10.1002/pam.20091

Hasan, S., & Kumar, A. (2019). *Digitization and Divergence: Online School Ratings and Segregation in America* [SSRN Scholarly Paper]. Social Science Research Network. https://doi.org/10.2139/ssrn.3265316

Hilton, M. (2016). *Prioritizing School Integration in ESSA State Implementation Plans* (Issue Brief No. 6). National Coalition on School Diversity. https://www.school-diversity.org/pdf/DiversityIssueBriefNo6.pdf

Holme, J. J. (2002). Buying homes, buying schools: School choice and the social construction of school quality. *Harvard Educational Review*, 72(2), 177 – 206. https://doi.org/10.17763/haer.72.2.u6272x676823788r

Houston, D. M., & Henig, J. R. (2021). *The 'Good' Schools: Academic Performance Data, School Choice, and Segregation* (EdWorking Paper No. 21 – 491). Annenberg Institute at Brown University. https://www.edworkingpapers.com/ai21-491

Hussey, O. C. (2020). *Measuring what Matters: Developing an Assessment Tool to Measure "Real Integration" in Schools* (Doctoral dissertation, Central Connecticut State University).

Hutt, E., & Polikoff, M. S. (2020). Toward a framework for public accountability in education reform. *Educational Researcher*, 49(7), 503 – 511. https://doi.org/10.3102/0013189X20931246

Johnson, R. C. (2019). *Children of the dream: Why school integration works*. Basic Books.

SOURCES CITED

Kho, A., Zimmer, R., & McEachin, A. (2022). A descriptive analysis of cream skimming and pushout in choice versus traditional public schools. *Education Finance and Policy*, 17(1), 160 – 187. https://doi.org/10.1162/edfp_a_00333

Knoester, M., & Au, W. (2017). Standardized testing and school segregation: Like tinder for fire? *Race Ethnicity and Education*, 20(1), 1 – 14. https://doi.org/10.1080/13613324.2015.1121474

Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Harvard University Press.

Koretz, D. (2017). *The testing charade.* University of Chicago Press.

Lee, J. (2008). Is test-driven external accountability effective? Synthesizing the evidence from cross-state causal-comparative and correlational studies. *Review of Educational Research*, 78(3), 608 – 644. https://doi.org/10.3102/0034654308324427

Lewis, A. E., & Diamond, J. B. (2015). *Despite the best intentions: How racial inequality thrives in good schools*. Oxford University Press.

Locke, J., Kang-Yi, C. D., Pellecchia, M., Marcus, S., Hadley, T., & Mandell, D. S. (2017). Ethnic disparities in school-based behavioral health service use for children with psychiatric disorders. *The Journal of School Health*, 87(1), 47 – 54. https://doi.org/10.1111/josh.12469

Marion, S. & Evans, C. (2021, August 20). Following their lead: Conversations about innovative assessment. Center for Assessment. https://www.nciea.org/blog/following-their-lead

Mexican American Legal Defense and Educational Fund. (2007, March 23). No Child Left Behind. https://www.maldef.org/2007/03/no-child-left-behind

Miner, M. A. (2020). *A Nation Rated? School Segregation and the Distribution of a School Resource: The Role of Accountability Ratings in Metropolitan Public Schools* [Ph.D. Dissertation]. University of Wisconsin Milwaukee.

Muñiz, R. & Barragán, S. (2022). Disrupting the Racialized Status Quo in Exam Schools?: Racial Equity and White Backlash in Boston Parent Coalition for Academic Excellence v. The School Committee of the City of Boston, *Fordham Urban Law Journal* 49(5), 1043-1090. https://ir.lawnet.fordham.edu/ulj/vol49/iss5/2

NAACP Legal Defense Fund. (2015, December 2). *LDF Statement on the Every Student Succeeds Act*. https://www.naacpldf.org/press-release/ldf-statement-on-the-every-student-succeeds-act

National Coalition on School Diversity. (2018) State of Integration 2018. Poverty and Race Research Action Council & the National Coalition on School Diversity. http://www.school-diversity.org/wp-content/uploads/2019/06/State-of-Integration_2018.pdf

National Coalition on School Diversity. (2020, May) Model State School Integration Policies, Policy Brief 11. Poverty and Race Research Action Council & the National Coalition on School Diversity. http://www.school-diversity.org/wp-content/uploads/NCSDPB11_Final.pdf

National Urban League. (2019). Standards of Equity & Excellence: A lens on ESSA state plans. National Urban League. http://ncos.iamempowered.com/pdf/NUL%20ESSA%20Executive%20Summary.pdf

Noonan, J., & Schneider, J. (2022). Beyond "good" and "bad": Disrupting dominant narratives about school quality. *Phi Delta Kappan*.

NYC Coalition for Educational Justice. (2019). *Chronically Absent: The Exclusion of People of Color from the NYC Elementary School Curriculum*. NYC Coalition for Educational Justice. http://www.nyccej.org/wp-content/uploads/2019/02/reportCEJ-Chronically-Absent-FINAL.pdf

Orfield, G., & Lee, C. (2007). *Historic Reversals, Accelerating Resegregation, and the Need for New Integration Strategies*. Civil Rights Project/ Proyecto Derechos Civiles. https://escholarship.org/content/qt8h02n114/qt8h02n114.pdf

SOURCES CITED

Orfield, G., Frankenberg, E., Ee, J., & Kucsera, J. (2014). *Brown at 60: Great progress, a long retreat and an uncertain future*. The Civil Rights Project/Proyecto Derechos Civiles. https://civilrightsproject.ucla.edu/research/k-12-education/integration-and-diversity/brown-at-60-great-progress-a-long-retreat-and-an-uncertain-future/Brown-at-60-051814.pdf

Pearman, F. A. (2021). Collective racial bias and the Black-White test score gap. *Race and Social Problems*. https://doi.org/10.1007/s12552-021-09347-y

Piazza, P. (2022). Back to the Future: Revising Old Critiques to Find a Culturally Sustaining Form of School Integration. In C. L. Ivery & J. A. Bassett (Eds.), *Detroit and the New Political Economy of Integration in Public Education* (pp. 127-148). Palgrave Macmillan. https://link.springer.com/chapter/10.1007/978-3-030-99796-0_10

Portz, J., & Beauchamp, N. (2020). Educational accountability and state ESSA plans. *Educational Policy*, 36(3), 717-747. https://journals.sagepub.com/doi/10.1177/0895904820917364

Potter, H., Lallinger, S., Burris, M., Kahlenberg, R., & Edwards, A. (2021). School Integration is Popular. We Can Make it More So. The Century Foundation. https://tcf.org/content/commentary/school-integration-is-popular-we-can-make-it-more-so

Quinn, D. M. (2020). Experimental effects of "achievement gap" news reporting on viewers' racial stereotypes, inequality explanations, and inequality prioritization. *Educational Researcher*, 49(7), 482 – 492. https://doi.org/10.3102/0013189X20932469

Raise Your Hand Texas. (2022, October). Measure What Matters: A report from the Measure What Matters Assessment and Accountability Council. Raise Your Hand Texas. https://www.raiseyourhandtexas.org/wp-content/uploads/2020/11/Measure-What-Matters-report.pdf

Ravitch, D., Forte, D., Moss, P., & Reville, P. (2022). Policy Dialogue: Twenty Years of Test-Based Accountability. *History of Education Quarterly*, 62(3), 337–352. https://doi.org/10.1017/heq.2022.19

reardon, s.f., Grewal, E. T., Kalogrides, D., & Greenberg, E. (2012). *Brown* fades: The end of court-ordered school desegregation and the resegregation of American public schools. *Journal of Policy Analysis and Management*, 31(4), 876 – 904. https://doi.org/10.1002/pam.21649

reardon, s.f., Robinson-Cimpian, J. P., & Weathers, E. S. (2015). Patterns and Trends in Racial/Ethnic and Socioeconomic Academic Achievement Gaps. In H. Ladd & M. Goertz (Eds.), *Handbook of Research in Education Finance and Policy* (pp. 491 – 509). Lawrence Erlbaum.

reardon, s.f., Kalogrides, D., & Shores, K. (2019). The geography of racial/ethnic test score gaps. *American Journal of Sociology*, 124(4), 1164 – 1221. https://doi.org/10.1086/700678

reardon, s.f. Weathers, E., Fahle, E., Jang, H., & Kalogrides, D. (2022). *Is Separate Still Unequal? New Evidence on School Segregation and Racial Academic Achievement Gaps* (No. 19 – 06; CEPA Working Paper). Stanford Center for Education Policy Analysis. https://cepa.stanford.edu/content/separate-still-unequal-new-evidence-school-segregation-and-racial-academic-achievement-gaps

Reid, K. S. (2005, August 30). "Civil Rights Groups Split Over NCLB." *Education Week*. https://www.edweek.org/policy-politics/civil-rights-groups-split-over-nclb/2005/08

Rogers, J. (2006). Forces of accountability? The power of poor parents in NCLB. Harvard *Educational Review*, 76(4), 611-641. https://doi.org/10.17763/haer.76.4.846v832864v51028

Ryan, J. E. (2004). The perverse incentives of the No Child Left Behind Act. *New York University Law Review*, 79(3), 932 – 989.

SOURCES CITED

School Diversity Advisory Group. (August, 2019). Making the Grade 2: New programs for better schools. https://steinhardt.nyu.edu/sites/default/files/2020-05/Making-the-Grade-II_0.pdf

Schneider, J. (2017). *Beyond test scores: A better way to measure school quality*. Harvard University Press.

Schneider, J., Carey, A. J., Piazza, P., & White, R. S. (2020). *School Integration in Massachusetts: Racial Diversity and State Accountability*. Beyond Test Scores Project and Center for Education and Civil Rights. https://cecr.ed.psu.edu/sites/default/files/Demography_Report_FINAL_7.24.20.pdf

Schneider, J., Jacobsen, R., White, R. S., & Gehlbach, H. (2018). The (mis)measure of schools: How data affect stakeholder knowledge and perceptions of quality. *Teachers College Record*, 120(5), 1 – 40. https://doi.org/10.1177/016146811812000507

Schneider, J., Noonan, J., Carey, A., Hersey, K., Marshall, J., McGovern, K., & Pavao, M. (2021a). Measuring School Quality Beyond Test Scores: A Toolkit. Beyond Test Scores Project & the Massachusetts Teachers' Association. https://static1.squarespace.com/static/60aea369b2b1517a8d2afb0c/t/60d3697f996b17546f8bfa5a/1624467840680/Measuring+School+Quality_A+Toolkit.pdf

Schneider, J., Noonan, J., White, R. S., Gagnon, D., & Carey, A. (2021b). Adding "student voice" to the mix: Perception surveys and state accountability systems. *AERA Open*, 7(1), 1 – 18. https://doi.org/10.1177/2332858421990729

Schneider, J., Piazza, P., White, R. S., & Carey, A. (2022). Student experience outcomes in racially integrated schools: Looking beyond test scores in six districts. *Education and Urban Society*, 54(3), 330 – 360. https://doi.org/10.1177/00131245211004569

Sorrells, A. (2022, October 17). Results from EdNC's survey on school performance grades. EducationNC. https://www.ednc.org/results-from-edncs-survey-on-school-performance-grades

Sosina, V. E., & Weathers, E. S. (2019). Pathways to inequality: Between-district segregation and racial disparities in school district expenditures. *AERA Open*, 5(3), p. 1-15. https://doi.org/10.1177/2332858419872445

Tienken, C. H., Colella, A., Angelillo, C., Fox, M., McCahill, K. R., & Wolfe, A. (2017). Predicting middle level state standardized test results using family and community demographic data. *RMLE Online, 40*(1), 1–13. https://www.tandfonline.com/doi/pdf/10.1080/19404476.2016.1252304

United States Government Accountability Office (GAO). (2016, April). K-12 Education: Better Use of Information Could Help Agencies Identify Disparities and Address Racial Discrimination, GAO-16-345. https://www.gao.gov/assets/gao-16-345.pdf

United States Government Accountability Office (GAO). (2022, June). *K-12 Education: Student Population Has Significantly Diversified, But Many Schools Remain Divided Along Racial, Ethnic, and Economic Lines*, GAO-22-104737. https://www.gao.gov/products/gao-22-104737

Vasquez Heilig, J., & Darling-Hammond, L. (2008). Accountability Texas-style: The progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy Analysis*, 30(2), 75 – 110. https://doi.org/10.3102/0162373708317689

Walker, V. S. (2018). The lost education of Horace Tate: Uncovering the hidden heroes who fought for justice in schools. The New Press.

Wells, A. S., & Holme, J. J. (2005). No Accountability for Diversity: Standardized Tests and the Demise of Racially Mixed Schools. In G. Orfield & J. C. Boger (Eds.), *School Resegregation: Must the South Turn Back?* (pp. 187 – 211). The University of North Carolina Press.