Measuring Opportunity: Redirecting Education Policy through Research

Ryan H. Pfleger
Newark Public Schools
Terri S. Wilson
Kevin G. Welner
University of Colorado Boulder
&
Charles Bibilos
United States

Abstract: While scholars develop research with clear implications for policy and practice, this work has been largely ineffective in influencing thought beyond the academy. In this article, we explore challenges researchers face in designing public scholarship to influence policy. To illustrate, we profile one such effort, the "Opportunity to Learn Index," a national project designed to compare opportunities to learn in all 50 states, and we detail challenges related to method, analysis and presentation. This type of effort, we conclude, asks researchers to embrace unique challenges and the inescapably political nature of the knowledge enterprise, especially when engaging with non-researcher audiences.
Keywords: opportunity to learn; education policy; public scholarship; research utilization

Medir la oportunidad: Redirigir la política educativa a través de la investigación
Resumen: Los académicos desarrollan investigaciones con claras implicaciones para la política y la práctica, pero este trabajo ha sido en gran medida ineficaz para influir en el pensamiento más allá de la academia. En este artículo, exploramos los desafíos que enfrentan los investigadores al diseñar becas públicas para influir en las políticas. Para ilustrar, describimos uno de esos esfuerzos, el "Índice de oportunidad de aprender", un proyecto nacional diseñado para comparar oportunidades de aprendizaje en los 50 estados, y detallamos los desafíos relacionados con el método, el análisis y la presentación. Este tipo de esfuerzo, concluimos, pide a los investigadores que acepten los desafíos únicos y la naturaleza política de la empresa del conocimiento, especialmente cuando interactúan con audiencias no investigadoras.
Palabras clave: oportunidad de aprender; política educativa; beca pública; uso de la investigación

Medir a oportunidade: Redirigir a política educativa para a investigação
Resumo: Os académicos desenvolveram investigações com claras implicaciones para a política e a prática, tendo este trabalho sido de grande importância para influir no planejamento mais de todo o mundo acadêmico. En este artículo, exploram os desafios que enfrentam as grandes e médias e as políticas para influir nas políticas. Para ilustrar, descrever unos de esos esfuzes, a "Índice de oportunidade de aprendizagem", um projeto nacional de planejamento para o plano de aprendizado nos 50 estados, e detectar os desafios relacionados com o método, a análise e a apresentação. Este tipo de esfriamento, concluísmo, pide sobre os desafios únicos e exclusivos da política da empresa do congestionamento, especialmente com interação com audiencias no investigadoras.
Palabras clave: oportunidade de aprender; política educativa; beca pública; uso da investigação

Measuring Opportunity: Redirecting Education Policy through Research

The test-based accountability approaches that have dominated U.S. education policy over the last two decades are grounded, in part, on the premise that measuring achievement will help drive school improvement. These approaches are also a testament to a maxim attributed to Peter Drucker: *What gets measured, gets managed.* As Linda Darling-Hammond (1992) observed, "whatever we choose to examine on a broad societal level is likely to influence our choices of strategies and our chances of solving the policy problem" (p. 238). So it has been with testing. As our data coffers began to overflow with test scores, we saw a corresponding shift in policies and practices.

Undeniably, these achievement measures are vital to our understandings of equity and outcomes. Recent achievement gap research has, for example, documented persistent test-score gaps between students from different racial groups and economic backgrounds, alongside disparities in graduation rates and other important measures of educational achievement (Berends & Peñaloza, 2010; Card & Rothstein, 2006; Clotfelter et al., 2006; Reardon, 2011).

Yet far less political attention has been focused on the disparities in *opportunity* that give rise to these disparate measurements of achievement (Carter & Welner, 2013; Duncan & Murnane, 2014; Murnane & Duncan, 2011). A long-standing—if largely unheeded—interdisciplinary research base highlights the need to address specific gaps in opportunity (Carter & Welner, 2013; Gamoran, 2001; Ladson-Billings, 2006; Milner, 2013; Rothstein, 2004, 2006). Indeed, researchers have called for

greater attention to the broad importance of opportunities to learn for decades, if not longer (e.g., Coleman et al., 1966; DuBois, 1949; Elmore & Fuhrman, 1995; Guiton & Oakes, 1995; Murnane & Duncan, 2011). These scholars have explained why policy and practice must reflect the best evidence about, for example, the significance of inequities in early childhood education, school funding, and economic conditions.

While researchers engage with one another in knowledge-building conversations around issues like opportunity gaps, a parallel policy conversation is also taking place among media, policy advocates, practitioners, policymakers and elected officials. The role of researchers in that second conversation depends in large part on how research is crafted and whether that research is brought into political conversations in deliberate and thoughtful ways (Ball, 1997; Henig, 2012; Oakes, 2017). Research products are not a *Field of Dreams* or a better mousetrap; if we merely build it, policymakers will not necessarily show up, nor will they beat a path to researchers' doors.

Research in the tradition of public scholarship, which may include action research, public sociology, service learning, community-based research and other ways to connect thought and action, attempts to engage broad publics through research collaborations, practices and products (Ball, 1997; Oakes, 2017). Our essay focuses on two central challenges of public scholarship, one of which focuses on the comprehensive research needed, and the second of which focuses on the necessary subjective and even political choices. The first challenge asks how the education research community can work cross-disciplinarily to develop a strong, comprehensive set of measures of educational opportunity and bring these to bear on current—and future—education policy conversations. The second challenge asks this same research community to embrace, or at least accept, the unique challenges and inescapably political nature of the knowledge enterprise, especially when engaging with non-researcher audiences.

This article presents a research-based analysis and essay, in the tradition of similar work from scholars such as Darling-Hammond (1992, 2004) and Henig (2012). As with these earlier efforts, our goal is to critically analyze the use of research and data alongside education politics. The word *essay* finds its roots in the Middle French *essai*, which itself is traced back to a Late Latin noun, *exagium*, meaning "act of weighing" (Merriam Webster online). Montaigne's *essais* in the late 17th century popularized the form, as an exploration of a topic or an attempt to test or prove a point. Accordingly, this work is an attempt to make sense of the challenges and possibilities of public scholarship. Our article draws on a specific ongoing policy-research effort (the "Opportunity to Learn Index" or OTLi) to make a larger argument about the opportunities and difficulties of policy-engaged research. We also consider how to negotiate the unique methodological difficulties posed by research projects that aim to influence policy directly.

Our goal is to describe and analyze our deliberative process that gave rise to the method, content, and purpose of the OTLi. We view this process as a case that might offer insights about the challenges and affordances of publically engaged research. In particular, in this article, we describe the methodological and moral puzzles we encountered in engaging in the OTLi project. In effect, this article offers a methodological reflection—grounded in our discussions in team meetings, as well as in the process of drafting and revising this essay—about the choices and compromises we faced in creating the OTLi, and in bringing it to bear on pressing issues in public policy. In this article we have resisted the temptation to offer a detailed description of the design and method of the OTLi itself. Our focus is on the lessons we learned through this project about the technical and normative challenges that can accompany public scholarship.

In the article's first section, we offer an overview of what we refer to as the "measuring-matters nexus" and explain the value of additional and improved measures of opportunity gaps. While some measures of opportunity exist (e.g., around health, housing and access to early

childhood education), other measures are less accessible, especially for a policy audience. We next describe the rise of comparative indices of student achievement and their influence on debates about opportunity. In the third section, we discuss our work to develop the Opportunity to Learn Index, a national index comparing opportunities to learn across all 50 states, and the unique methodological challenges of this approach. In the final section, we offer some principles and guidance for thinking through the difficulties of engaging in public scholarship and policy research designed to inform and redirect educational opportunity.

Measuring Achievement and Opportunity

The academic enterprise is largely one of knowledge building. In some cases, scholars labor to build evidence and truth and to advance science through basic research. In other cases, and often in education, scholars work to find solutions to pressing issues of quality and equality. These solutions, however, only become broadly accessible when they move from the knowledge-building conversation to the policy-making conversation. When researchers limit themselves to a largely insular conversation, this interchange is much less likely to occur.

While knowledge-building conversations can productively surface or generate disagreement, they can also provide clear direction. An example—one that we have been focused on in our own work—is found in the cross-disciplinary knowledge-building conversation that has identified multiple dimensions of opportunity gaps (Carter & Welner, 2013; Duncan & Murnane, 2014; Murnane & Duncan, 2011). These findings draw, for example, on expertise about how to teach emerging bilinguals and children with special needs, as well as how to teach science, literacy, music, arts, history, civics, and mathematics; on expertise in foundational disciplines such as anthropology, sociology, philosophy and history; and on specialists in educational psychology and the learning sciences. Knowledge-building conversations are also informed by specialists in research methods, in leadership, and in the preparation and induction of teachers and counselors, as well as by scholars who examine political and legal obstacles to enacting and implementing beneficial policies.

The knowledge base about opportunity gaps will not, however, have much influence on policy and practice unless two other pieces are also in place. First, researchers interested in advancing public scholarship must create *usable information* regarding opportunity gaps.[1] Second, and more controversially, such researchers must join with others—including allied advocacy and organizing groups—in *directly advocating* for the use of that information. Little is accomplished by the mountain of broad, vague calls for policymakers to heed high-quality scholarship. Researchers pursuing public scholarship must also engage in the distinctly political work of influencing and shaping policy. Here, Henig (2012) reminds us that data, politics, and policymaking are necessarily intertwined. Politics, he argues, is not something to overcome; it should be "one of a series of challenges to be acknowledged and taken into account" by researchers (p. 3).

The potential impact of this engagement is illustrated by the political and rhetorical salience of the achievement gap framework and its connection to the politics and practices of accountability. One reason the achievement gap framework is so powerful is that it offers easily digestible data

---

[1] Not all research should directly aim to influence policy and practice. Many studies in education generate valuable knowledge without seeking directly to influence policy conversations. Accordingly, our framework and guidance is only for research that does straightforwardly intend to shape policy. This article thus includes phrases such as "policy relevance" and "publically engaged" as shorthand to describe some goals of researchers. Exploring the complexity of these goals is not the focus of this essay. For more on the debates and complexities about research goals such as relevance and engagement, particularly in the context of sociology, see Nichols (2007).

about outcomes, in the form of indices, assessments, and other tools that enable comparisons between states, districts, schools, teachers and students. The drive to measure achievement—and, in particular, the achievement gap between middle-class White students and lower-income students of color—has played an increasingly influential role in framing education policy and narrowing educational practice (McNeil, 2000; Valenzuela, 2005; Wiley & Wright, 2004). Moreover, the very idea of an achievement gap has shaped public conversations and policy priorities about education (Ladson-Billings, 2006).

A few decades ago, this focus on achievement gaps and performance standards was not a foregone conclusion. Welner (2010, p. 92) described this moment as follows:

> Back in the late-1980s, advocates of equal educational opportunity developed the concept of 'opportunity to learn' (OTL). OTL focused on resources—on inputs. Students provided with OTL would have access to high-level schooling opportunities, allowing them to meet performance and content standards and providing them with good career opportunities (Oakes, 1989; see also Darling-Hammond, 1994; Stevens & Grymes, 1993). One of the natural corollaries to this argument was that students should not be held to high performance standards unless their schools have met equally high OTL standards. . . .
>
> But the focus on inputs—the OTL approach—never gained enough favor among political leaders, who opted instead to put their faith in the idea that the process of measuring outcomes and holding schools accountable for meeting certain outcome objectives will itself drive improved practices. . . . Instead of providing better resources, the outcome approach seemed designed to drive better use of existing resources. These standards-accountability reforms, ultimately epitomized by the No Child Left Behind Act (NCLB), rely upon content and performance standards plus high-stakes testing. In theory, such standards and tests (aligned with inputs like teacher preparation and textbooks) will drive the educational establishment to ensure a quality education for every child (Smith & O'Day, 1991).

This triumph of an achievement framing over an opportunity framing set the nation on its current path, some benefits of which should be acknowledged.[2] Perhaps most importantly, NCLB challenged what President Bush (2004) called the "soft bigotry of low expectations," including hiding low test scores—and the students with those scores—beneath aggregate data in otherwise high-scoring schools (Losen, 2003).

In doing so, however, the NCLB-era unleashed an avalanche of test-score data, and these data have changed the purpose, content, process, and politics of education. To be clear, test-score data do not, by themselves, determine their own use. Use is guided by context, judgment and policy. The effects of achievement measures, for instance, depend on whether they are deployed in summative or formative ways, the extent to which they are used in high-stakes environments, whether they are focussed on specific content domains (e.g., reading and math), and how they are

---

[2] Several studies have evaluated reforms that foreground the use of achievement data, including: performance measurement (Propper & Wilson, 2003; Verbeeten, 2008), data-driven district initiatives that focus on benchmark assessments and data consultants (Carlson, Borman, & Robinson, 2011; Slavin, Cheung, Holmes, Madden, & Chamberlain, 2013); assessing and reassessing student reading comprehension, creating reports and employing data-coaches (Quint, Sepanik, & Smith, 2008); benchmark and formative assessment reforms (Konstantopoulos, Miller, & van der Ploeg, 2013), and high-stakes testing and test-driven accountability (Jacob, 2005; Lee, 2008). These studies, several of them experimental, came to mixed conclusions about the academic effects of focusing on achievement data.

connected to policies of standardization and accountability. Nonetheless, the availability of test-score data shapes broad research and policy agendas. The rise of the Data Scientist and the prestige of Big Data (Boyd & Crawford, 2012), coupled with the proliferation of programs like the Strategic Data Project (of which one author of this article was a fellow), suggest that data are significant touchpoints, if not origins, of changes in education policy and practice.

The move towards high-stakes testing, argues Gunzenhauser (2003), may even have changed the several factors that constitute the multi-faceted enterprise of schooling. For example, NCLB and test-based accountability structures that turn to a menu of centralized and corporate reform strategies undermined democratic local governance of schools (Au, 2010; Howe & Meens, 2012). For example, Howe and Meens wrote, "While local school systems are accountable for performance, performance itself is measured in terms of scores on standards-based tests developed and enforced from afar" (p. 3). Instead of a local community choosing what to measure, the criteria for good schooling are developed at a more distant level of government that makes personal kinds of democratic deliberation more difficult. Additionally, test scores are sometimes used by parents to choose the best schools for their children. This "voting with your feet" and individualized competition may substitute for democratic, community-level deliberation about the goals and governance of schooling. As a consequence of sanctions that stem from test-based accountability, "decision-making power is taken out of the hands of the communities most affected," which is at odds with democratic local governance (Howe & Meens, 2012, p. 9).

Test-based accountability, and especially test preparation activities that have arisen from it, also change the nature of instruction and are associated with math teaching that is of lower quality and less rigorous (Blazar & Pollard, 2017). Instruction is modified by the interim assessment results that have now become a dominant feature of classrooms (Clune & White, 2008). Moving from a broad to a narrow curriculum (Au, 2011; Berliner, 2011), and even towards a scripted curriculum (Wright, 2002), changes educational content and pedagogy. Accountability through testing often leads to focusing interventions on "bubble kids" (Booher-Jennings, 2005) and has prompted cheating and "gaming" to improve results (Thompson, 2011). Achievement data also prop up problematic practices such as ability tracking and grade retention (Shepard, 1991).

The nation's test-score focus arises, in part, out of a political process that has increasingly held schools accountable for tested achievement. Political actions have created a situation where schools (the presumed "cause") are held accountable for achievement (the "effect"). However, achievement and its causes are complex. First, achievement depends on school goals (i.e., achievement is normally measured in terms of literacy and math, but it could include democratic dispositions, music proficiency, social skills, physical health, and more). This "accountable-for-what" problem cannot be resolved by reference to the usual math and literacy test-score achievement data. Second, many kinds of academic and other desirable child outcomes are functions of disparate actions. These are linked to the choices, skills and behaviors of teachers, schools and students; to funding decisions of politicians; and to larger societal inequalities related to factors such as racism and economic inequality (Carter & Welner, 2013). In short, the relationship between cause and effect in education is not as straightforward as is acknowledged in test-based accountability reforms.

Within the last few decades, accountability based in standards of achievement and measures of achievement became the "normal" way to view education reform (McNeil, 2000). In a study of the political machinations that led to the development of standards and testing reforms associated with *No Child Left Behind*, McGuinn (2006) argued that an "accountability regime" replaced the previous "equity regime." The focus of the equity regime, which included the original Elementary and Secondary Education Act, was on providing additional funding to schools with students who were poor, and it included the promotion of racial integration. The accountability regime shifted the

focus to measured student achievement, including a focus on test-score gaps across race and socioeconomic position.

Moving away from providing financial resources changes education politics because it shifts the burden from communities with resources (to provide resources) to marginalized communities (to provide higher scores). Similarly, focusing on measures of student achievement, instead of measures of inequities in opportunities outside of school, shifts the social burden of equity from people and institutions who are politically, economically, and socially powerful to schools, teachers, and students. All these changes are inextricably tied to the decisions during the last decade of the twentieth century to set aside the call for opportunity to learn standards and to instead embrace performance standards tied to testing. The policy connection between what data we collect and what we then address—the "measuring-matters nexus"—may not be desirable, but it is undeniable.

To be clear, outcome data focused on achievement and attainment are a valuable part of a healthy evaluative feedback loop. The concern here is one of balance. The single-minded focus on test scores has often been counter-productive for broader concerns about equity (Welner & Mathis, 2015). Might balance be restored if researchers develop comparably useful information regarding opportunity gaps? Would policy follow? In the next section, we illustrate the contrast between these two different frameworks by analyzing an increasingly common and influential way to present data for policy audiences: indices and comparative measures of academic performance.

The Achievement Gap and the Influence of Indices

As noted above, accountability measures and indicators have grown in scope and influence over the last 30 years (McDermott, 2011; McDonnell, 2004), and standardized assessment data play a particularly dominant role in education policy (Planty, 2010). These data are often organized into composite indices and presented in a wide variety of forms, including letter grades, percentages, or rankings of schools, districts or states (Darling-Hammond, 1992). Indices have become ubiquitous throughout the education system, and many were created as part of accountability regimes (Howe & Murray, 2015). For example, the state of Colorado sorts schools into four color-coded categories according to its School Performance Framework, which includes measures of academic achievement, academic growth, academic gaps, and graduation and dropout rates (Colorado Department of Education, 2017). The School Performance Framework index is then used to inform administrative decisions at the district and state level, including the assignment of turnaround status for schools that consistently score lowest in four categories (Colorado Department of Education, 2015).

Education indices have also gained substantial prominence and influence outside the realm of government-sanctioned accountability frameworks. Advocacy groups, magazine publishers, and websites have released numerous indices and rankings in recent years. Perhaps the best-known example is *U.S. News and World Report*'s annual college rankings, which were first published in 1983 and have since been joined by rankings from *Atlantic Monthly* and *The Wall Street Journal*, among others (Hazelkorn, 2015). The commercial and cultural success of the *U.S. News and World Report* college rankings have spawned a cottage industry of similar efforts, including rankings of preschools, high schools, graduate programs, and state education systems on a wide variety of metrics (see, e.g., Hazelkorn, 2015). Many of these indices and rankings are simply intended to help parents and students make informed educational choices, although others use rankings and grades as advocacy tools to advance particular policy agendas. Unfortunately, very few indices focus on educational resources or are based on high-quality research (for exceptions, see Baker et al., 2015; Barnett et al., 2017).

Even under the best of circumstances, indices often suffer from a wide range of methodological weaknesses. First, rankings and indices require researchers to determine schemes for weighting components of the index, and these judgments are generally only partly justified (see analysis in Camilli & Firestone, 1999). For example, in creating Colorado's School Performance Framework, state officials chose to give test-score data a substantially more prominent role in the index than dropout rates (Colorado Department of Education, 2017). While all research involves value judgments (Howe, 2003), indices rarely make these value-laden decisions clear and transparent (Camilli & Firestone, 1999).

Combining metrics that appear on different scales similarly requires subjective choices. For example, *Education Week*'s Quality Counts index includes measures of school finance and measures of student achievement that are on entirely different scales; combining these measures into a single, interpretable number requires a mathematical transformation. For each measure used in its index, *Education Week* chose to set the high score to 100 and then calculate the other states' scores as a percentage of the highest score. Alternative aggregation methods might include standardization or summation of rankings, but each of these options presents its own set of pitfalls and challenges (Nardo, Saisana, Saltelli & Tarantola, 2005).

Even when weights and transformations are chosen thoughtfully, rankings and indices inevitably present simplifications of educational realities and obscure nuances and flaws in the underlying data. In an analysis of school report cards used in state accountability systems, Howe and Murray (2015) found that the common use of a single summary A–F letter grade, although often intended to empower stakeholders by providing simple measures of school quality, actually creates confusion. The researchers concluded that a single composite score is a "dubious proposition . . . because it is not clear what a single grade can mean across a diverse array of criteria" such as achievement, class offerings, "college and career readiness," and many other processes and outcomes (Howe & Murray, 2015, p. 5).

Despite these limitations, educational indices and rankings are seductive in their simplicity and often have a substantial impact on decision-making. The powerful impact is readily apparent when indices are used explicitly in accountability decisions, such as Colorado's use of the School Performance Framework to assign "turnaround" status to schools. In the case of indices that are designed to inform public opinion, the apparent impact on policy and on individuals' educational choices can again be surprisingly powerful. For example, Luca and Smith (2013) have shown that applications to colleges increase by nearly 1% for every one rank improvement in the *U.S. News & World Report*'s annual college guide. Similarly, Jin and Whalley (2007) found that state universities that received coverage in the *U.S. News & World Report* rankings subsequently received substantially more state funding than those that did not appear in the rankings guide. Although other educational indices may attract less attention than the *U.S. News & World Report* rankings, it is not far-fetched to imagine that any index that is widely read or widely covered in the media will ultimately affect education policy, even if the index is based on dubious methodology or flawed underlying data (see Welner, 2015).

Since education rankings and indices likely influence education policy and public opinion, it seems likely that a carefully crafted opportunity-gap index could raise the profile of those measures. Again, however, simply building such an index and making it publicly available will likely do little. The mere presence of an index that highlights opportunity gaps does not, by itself, guarantee that the index would have an impact on education policy. Researchers, for instance, developed school-segregation indices several decades ago (see Clotfelter, 1998), but the problem has actually received decreased policy attention (Orfield & Frankenberg, 2014). This is partly because education is a national political issue with a cacophony of voices and interest groups that often drown out

conventional education research (DeBray-Pelot & McGuinn, 2009). Further, advocacy think tanks and advocacy-based research have played an increasingly large role in directing education policy (Scott, Lubienski, & DeBray-Pelot, 2009), and those efforts have generally pointed attention away from concerns about racial segregation.

In this environment, indices—even those, like the measures of segregation previously mentioned, that are important and of high quality—and other policy-relevant research can be shoved aside in public policy debates. Researchers hoping to bring their work into these conversations may need to design measures with an end goal in mind: their potential use and reception in public policy debates in education. Moreover, as we discuss later in this article, researchers engaged in public scholarship may find it useful to join with non-researchers in coalitions to more directly communicate these data and analyses to policy-makers and practitioners, in formats that resonate with their concerns.

Policy-relevant and applied research, however, raises difficult methodological challenges. How might researchers design measures that will inform—and even directly influence—policy debates, in ways that are relevant as well as rigorous? This is not an easy path to tread, and it involves a unique series of methodological choices and compromises. Moreover, there is little guidance for how researchers might negotiate such difficult steps (for an exception, see Willinsky, 2001). The following discussion of our work to develop an "Opportunity to Learn Index" explores that process. We do not offer this project as an exemplar or best-practice, but as a window into the difficult methodological choices—and necessary compromises—involved in translating research to inform current policy conversations around educational opportunity and inequality.

The Opportunity to Learn Index

The "Opportunity to Learn Index" (OTLi) is part of a larger project at the National Education Policy Center focused on opportunity-to-learn issues.[3] This section draws on that OTLi work to analyze key challenges and consider how these challenges might offer insight for others taking on this type of research. The OTLi is grounded in a collection of evidence-based essays from a diverse set of scholars, collectively describing the opportunities denied many American children (Carter & Welner, 2013). Still in development, the national index compares opportunities to learn in all 50 states along approximately 15 dimensions, some of which, although important, are not currently measured (and may not be measurable): racial segregation, child health, family social class, school finance, access to preschool education, standards and accountability, school choice, school cultures, needs of language minorities, discipline, tracking and segregation, teacher quality, more and better learning time, safe school environments, and opportunities for democratic participation. Within these dimensions are 70 specific indicators. The index is essentially a report card evaluating the provision of opportunities to learn.

In this section, we focus on the methodological decisions made in creating the index, rather than on actual rankings or results. We do so with an eye to similar challenges that may confront other researchers seeking to engage policymakers in similar ways. Specifically, we focus on five overlapping challenges we negotiated in building the OTLi: (1) defining core concepts; (2) identifying measures of opportunity from enormous bodies of literature; (3) communicating the relative significance of dimensions; (4) combining measures, and (5) presenting results. These challenges required our research team to balance our desire to portray important nuances in the existing research base against the need to create short summaries and simplified measures that would

---

[3] The NEPC "Schools of Opportunity" recognition project for public high schools, found at schoolsofopportunity.org, is the most prominent of these projects.

be attractive and usable to policymakers, the media, and the public. Again, we do not present these challenges to justify the choices we made in the OTLi, but rather to provide examples of the types of challenges researchers may face in redirecting education policy through research.[4]

Challenge 1: Defining Core Concepts

The concept of opportunity to learn has been explored for decades (see Guiton & Oakes, 1995; Oakes, 1989), but it was surprisingly difficult for our research team to create a precise, working definition that could be operationalized into a quantitative index. To illustrate, consider the dimension of economic opportunity. Clearly, the economic background and status of a student's family plays an enormous role in his or her learning opportunities, and countless studies show a strong relationship between family income and academic outcomes in the United States (e.g., Coleman et al., 1966; Collins, 2009; Murnane & Duncan, 2011). But what, exactly, does it mean to include "family income" in the OTLi, and how precisely does income impact opportunity? Moreover, if income is included, should wealth—which is even more unevenly distributed (Piketty, 2014) and educationally significant (Orr, 2003)—also be included?

If we were to include a state's median income or wealth in the index, the project would imply that the *level* of economic inputs determines opportunity. In a society with low economic inequality (a category that the United States does not fit within), economic level indicators would illuminate a significant portion of the relationship between economic resources and educational opportunity. A second approach might be to use a measure of economic inequality, such as a Gini coefficient (see Piketty, 2013, for a discussion of the merits of the Gini coefficient). Including measures of inequality would suggest that the *inequalities themselves* are a hindrance to educational opportunity. A third approach would be to use economic policy measures, such as the fairness of a state's tax structure or the existence of state-level living wage laws. Including these measures in the index would suggest that *economic policy* affects educational opportunity. Finally, a fourth approach would be to (somehow) include measures of the historical, systemic, and cultural underpinnings of economic inequities, an approach that would suggest that institutionalized racism or neoliberal economic policies have contributed over time to an increasingly polarized and inequitable society (Ladson-Billings, 2006). Any combination of these four approaches could be defensible, depending on how one conceptualizes the relevant causes of economic opportunity.

Put more simply, it is unclear whether "opportunity" should be conceptualized as a condition—measured by a level indicator such as family income—or as the factors that help shape that condition, such as the Earned Income Tax Credit, which influence the economic resources available for families (Marr, Huang, Sherman, & DeBot, 2015). If we had chosen to define "opportunity" solely as a condition, we would implicitly de-emphasize the policies and systemic factors that lead to inequities. If we had chosen to define opportunity solely in terms of the policies that create inequities, then we would be forced to make some strong assumptions about exactly which policies cause opportunity gaps. We ultimately chose a balance of both, but we chose to include policy measures only when their impact on educational opportunities was strongly supported by existing research. The first challenge of policy-engaged research, in sum, is defining and presenting core concepts in ways that are both politically relevant and grounded in research.

---

[4] The idea of the index project was strategic and political. We chose to pursue the index in order to capture public interest, invite reporting, and encourage states to think about how they stack up against one another. As explained in this article, we concluded that an index would bring attention to opportunity-to-learn issues. Our initial decision to engage in this way then led us to challenges and uneasy compromises: to simplify for the sake of clarity, to design measures that would enable state comparisons even when those measures were incomplete, and so on.

Challenge 2: Identifying Measures of Opportunity from Enormous Bodies of Literature

The OTLi is organized, on a basic level, around summaries of 15 broad areas of research literature. While we made use of the literature surveyed in an edited volume focused on opportunity to learn issues (Carter & Welner, 2013), we also surveyed a wide range of additional research in each dimension, attempting to identify specific indicators and available state-level data. This was one of the most persistent challenges of the OTLi. There are multiple, complex dimensions of educational opportunity (e.g., consider the many and various components of child health). The literature within each of these dimensions is also enormous. Some dimensions, like social class inequality, have been studied for many decades and through multiple disciplinary angles. In addition, specific indicators often have generated their own voluminous literature. Food security, for instance, while just one indicator within one dimension (child health) in our overall index, has itself been the subject of vast amounts of research (see Coleman-Jensen, McFall & Nord, 2013).

Accordingly, we decided early on to identify authoritative reviews of the existing literature. These reviews existed in only a few areas but, when available, allowed us to rely on expert overviews of complex fields. For example, the National Institute for Early Education Research publishes an annual index of early childhood education with multiple measures of access and quality (e.g., Barnett et al., 2017). A second way we confronted the overwhelming research-summary task was to prioritize transparency in the choices we made. Given the multiple challenges posed by inevitably incomplete, inconclusive and conflicting research in many areas, we sought to be fully transparent in reporting key decisions regarding index components. So whether we relied on expert summaries of research, expert indices, or on our own reading of a research area, we sought to clearly disclose the sources of evidence used to draw those conclusions.

A challenge specific to creating an index is identifying defensible indicators. We began with the research about the causes of opportunity gaps and then considered possible measures, ultimately compiling 70 such indicators. Sometimes this process went smoothly, as when creating the dimensions for school finance and preschool, where the research base is robust and other researchers had already compiled defensible measures. Very often, however, we encountered difficulties. Identifying something that *should* be measured is very different than identifying the measurement itself. For example, there is ample evidence that placement of students into stratified "tracks" within schools is detrimental to student learning (Oakes, 2005), but no dataset could be identified that provides reliable and useful state-level measures of student tracking within schools. School choice is another example. While research suggests that charter schools stratify opportunities when they deny equitable access (Scott & Wells, 2013), there is no set of easily counted or categorized differences across the state laws enabling charter schools, and several states have no charter school law at all. The OTLi therefore only *discusses* the research and types of possible indicators of equitable access to charter schools; it omits quantitative indicators regarding charter schools or other forms of school choice from the index. Similarly, a positive, inclusive school culture is a necessary precondition for equitable learning opportunities (Carter, 2013), but quantitative indicators about school culture, especially across states, are elusive.

Creating effective policy-engaged research also calls upon researchers to accept the inevitability of making judgments grounded in expertise and of ultimately making use of incomplete evidence. Consider the fact that, by itself, the National Survey on Child Health includes 70 indicators (The Child and Adolescent Health Measurement Initiative, 2013), and numerous other health indices are available. Should the child health dimension of the OTLi attempt to include all of these? We explored much of the existing research on the links between, for example, oral health and student learning, as well as the relationship between untreated vision problems and student learning. On the one hand, it stands to reason that a student who cannot see the chalkboard or a student who

is distracted by a severe toothache would struggle to learn; on the other hand, it is nearly impossible to define exactly *how* important oral health or vision care are positioned within the scheme of a child's overall health. Are these more important than, say, a condition like malnutrition or a policy that influences access to health insurance? Any effort to pare down indicators is fraught with such decisions. Even with a commitment to base our decisions on the highest-quality research, we inevitably had to select measures based on a mixture of incomplete evidence and our own judgment.

Such situations, which were common, presented us with: (a) important, persuasive research suggesting an indicator that should be included in the OTLi; but (b) no useful existing measure. If our goal was solely to create an internally valid quantitative index, we might simply have chosen to ignore the dimensions that lacked high-quality indicators. However, this approach would have undermined the work's construct validity and external validity as well as our broader goal of conducting engaged research with the potential to inspire a more inclusive policy debate about educational opportunity. Omitting a dimension from our index would not promote its inclusion in the broader conversation. As a result, whenever possible, we chose to include the best available indicators for each of our dimensions, while also expressly noting the flaws and gaps in available information. Our belief is that awareness might inspire policymakers and researchers to develop better measures and improve data collection. To further this end, we made a point of stressing that our composite index scores are not comprehensive. We were also deliberate in including dimensions of the opportunity gap that do not yet have state-by-state data. We decided that narratives about components without state-by-state data are as important as the numbers. This second challenge, in sum, concerns distilling complex literatures in ways that point to evidence-based policies, while being transparent about lacunae.

Challenge 3: Conveying Relative Significance

Once we identified available sources of data, our team still faced difficult questions about how to organize and weight various measures within each dimension, as well as how important each dimension would be within an overall composite score. The general importance of certain indicators can be fairly easy to grasp conceptually. For example, in the child-health dimension, it is not hard to understand that lack of access to food or medical care might prevent learning. But compromises were necessary in order to translate each dimension into numbers that illustrate potential variation across states.

The OTLi includes 15 primary dimensions, and any potential weighting scheme for these dimensions inevitably involves difficult choices and judgments about their relative importance. Weighting them equally sounds "fair," but it ignores the fact that some dimensions almost certainly have a larger effect on learning opportunities than others. We could not, for example, justify a weighting scheme where socioeconomic status determines only 1/15th of a child's opportunity to learn. However, any other weighting scheme again involves difficult judgments that others may justifiably disagree with. Our solution was to explain our decisions clearly, and to design interactive mechanisms on the OTLi website that would allow readers to disagree and set their own weightings for each dimension.

These decision-making dilemmas have much in common with the research compromises made in any study. Our proxy choices are, for example, comparable in some ways to those faced regularly by quantitative researchers in their specification of models—imperfect representations of the actual phenomenon. Similarly, qualitative researchers (e.g., Miles & Huberman, 1994) have drawn on Simon's (1957) concept of "satisficing" to describe criteria for making decisions in research design and data collection. Consider as well the choices made in program evaluations, concerning questions, data, and analytic approach that often depend on the intended use by the

intended user (Patton, 2008). According to Cronbach (1982), evaluations are usefully approached not as uncompromising "science": "Designing an evaluative investigation is an art. The design must be chosen afresh in each new undertaking, and the choices to be made are innumerable. Each feature of a design offers particular advantages and entails particular sacrifices" (p. 1).

Guided by the principles of defensibility and transparency, our decisions about weighting and organizing data for this project used three primary criteria. First, we prioritized data that depicted the broadest set of factors consistent with research. For example, a diverse set of measures was included in the child health dimension, including health insurance, food insecurity, mental health care, oral health, and asthma. Bodies of research link these factors independently to important outcomes. The best data sources also included clear and comprehensive details of the concept measured, collection processes, and generalizability within and across raced and classed groups. Second, we sought data sources that demonstrated how a given dimension varied with respect to raced, classed, and language groups. Third, we assigned weights in an ongoing, iterative process, guided by the strength and availability of the evidence in a given area, as well as by the overall importance of a dimension in the reviewed research.

For example, research has established a strong causal relationship between preschool participation and student outcomes, and it has also clearly defined some of the specific elements of "high-quality" preschool programs (Barnett & Lamy, 2013). Furthermore, the National Institute for Early Education Research provides useful state-level data that could be used in an index (e.g., Barnett et al., 2017). In contrast, consider the research on blood lead levels, which is just as conclusive, if not more so (U.S. Department of Health, 2007). The effect sizes are just as large, if not larger, and we see a disparate impact on learning that varies across race and class (Berliner, 2009). But we discovered that state-level comparative data were not available. The best data came from a single federally administered dataset (Centers for Disease Control and Prevention, n.d.) that covered most states, but—on close reading—the data were not comparable across states. Thus, blood lead levels are not included as a quantitative component of the OTLi. While we include a narrative section about the importance of lead, the lack of a reliable and comparative measure compels an unfortunate but necessary omission—one that points again to the need for researchers (and governmental authorities) to fill the void.

These decisions, as Cronbach (1982) cautions with regard to evaluation, involve both art and judgment. In all cases, we attempted to be as transparent as possible about our justifications for using particular measures and for how such measures are weighted. However, our uneasiness in this process also underscored the importance of designing an online version of the index where users are invited to select alternative weighting schemes. While we assign provisional weights to more than 70 indicators across 15 dimensions, these decisions offer readers only a baseline interpretation, which can be revised to reflect differing values, local conditions and norms. In this way, we sought to make the index relevant to a broad set of stakeholders (some of whom might specialize in just one of the dimensions), while providing provisional weights.

Challenge 4: Combining Measures

A related issue concerns how to combine data that are on different scales and that use different distributional forms. Whenever a composite index includes indicators that are measured on different scales, the mathematical transformation used to combine the indicators will leave researchers open to criticism. For example, indicators may be transformed into percentages, percentiles, individual rankings, or z-scores before they are averaged to create a composite index, and each of these techniques has its own strengths and weaknesses (Nardo et al., 2005).

A related flaw in combining measures is that our dimensions surely overlap with each other. Even though the OTLi is not a regression model and is best understood as a heuristic device, the index is complicated by the fact that the addition of some dimensions effectively "double-count" overlapping elements of opportunity gaps. Consider the overlapping components of child health (partially illustrated in Figure 1). The basic research convincingly suggests that child health has a strong effect on student learning. But within this dimension we included a measure for the availability of health insurance, which surely overlaps with another measure, mental health, since health insurance often supports access to mental health services. Similarly, within the separate dimension of early childhood education, there are measures of health that include screenings and referrals for vision and hearing along with the provision of at least one meal per day. A third dimension, school-resources, is concerned primarily with adequacy and inequality in school finance. But it is certainly associated with measures of schools' ability to provide health services and nutrition. Furthermore, the social class inequality dimension has measures of income and wealth that are important, in part, because they are associated with multiple facets of child health.
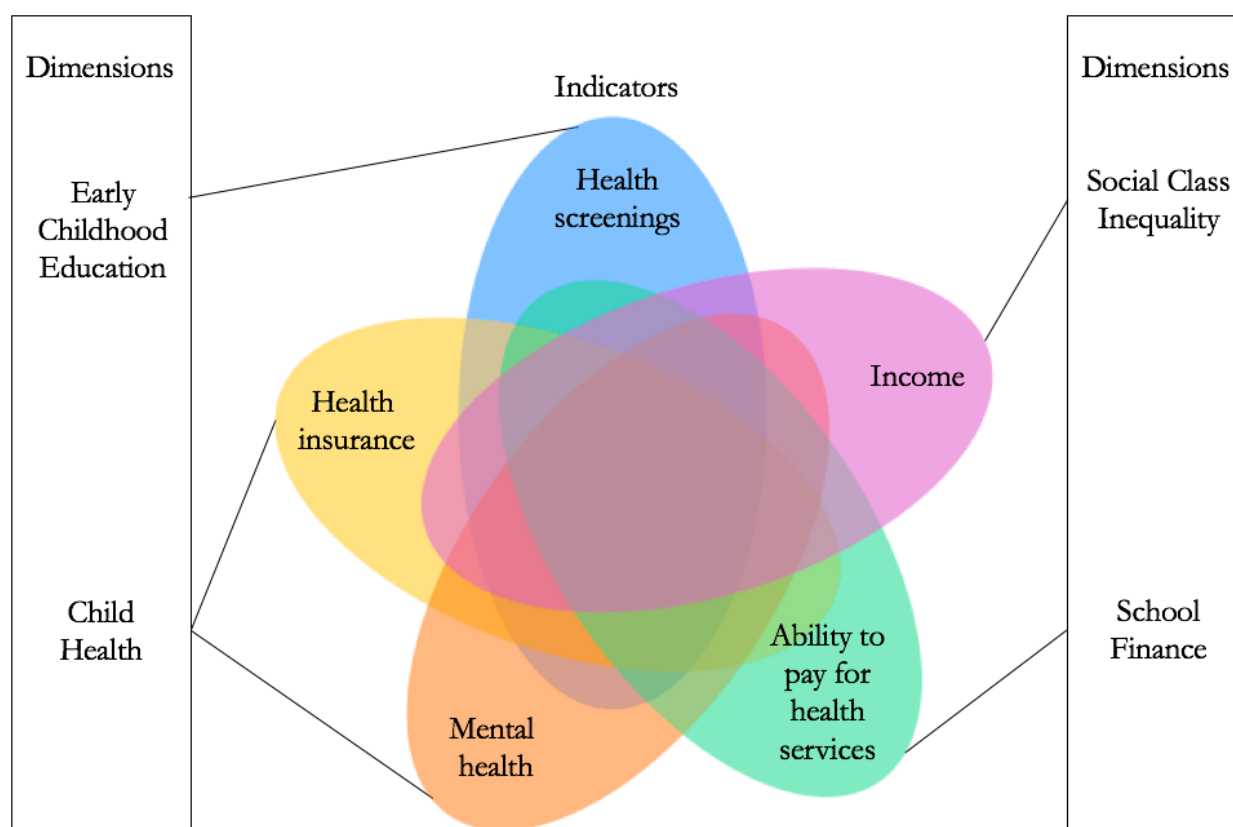
Figure 1. Example of indicators and dimensions that overlap.

Although the resulting overlap creates the likelihood of double counting (and should be avoided in causal research designs), research consistently finds that both income and health have independent effects on desirable educational outcomes. Adding to the complications of overlapping indicators, policy solutions may impact child health quite directly—through, say, school-based health clinics (see, e.g., McNall, Lichty, & Mavis, 2010)—and more indirectly through, say, income

supports like the Earned Income Tax Credit (see, e.g., Hoynes, Miller, & Simon, 2015). Again, these overlapping indicators are problematic for researchers who aim to discern discrete causal relationships—for instance, ascertaining the unique causal impact of health on learning needs to take into account the simultaneous role of income on learning and health. Such causal research is valuable; indeed, most decisions to include a dimension within the OTLi are based on just this kind of research.

Our approach to dealing with overlapping measures was to err on the side of simplicity and transparency while using data transformations that give reasonable composites. Although the various components may not measure truly distinct aspects of opportunity, presenting overlapping measures offers policy makers multiple ways to understand and potentially rectify inequities. We also pursued consequential validity (Carter & Welner, 2013), mindful that the measures are in service of an accountability and improvement goal: focusing attention on closing opportunity gaps. Again, the pragmatic nature of this type of translational work calls upon researchers to make a series of uncomfortable choices. The only way to avoid the choices is to eschew the work itself, along with its potential to bring attention to the underlying compelling research.

Challenge 5: Presenting the Results

The standard format of research reports—e.g., long journal articles—often works against its ability to influence policy. Yet aiming to address this concern also poses some challenges. How can researchers distill and present results in ways that are concise, understandable and engaging? What gets left out, and what impact may these choices have on the trustworthiness and transparency of publically engaged scholarship?

In our project, we sought to balance the value of concision against other standards for scholarly research (see American Educational Research Association, 2006). We designed the OTLi to use a polished format that did not follow prevailing norms of scholarly research. We also designed two versions of the OTLi product—a shorter version, along with a lengthy and nuanced report available for download. Further, the online version, including the user-adjustable weights noted earlier, promotes engagement with the index but also encourages reflection on the difficult choices made in any index process. This multimodal presentation format will allow for transparency about key operational decisions and will serve as an entry point for user-customization and public engagement. These decisions add expense as well as their own form of unwieldiness. Balancing standards and norms of academic research with presentation formats designed to effectively engage policy debates will remain a challenge for researchers aiming to make research relevant.

Policymaking takes place within a context that is shaped by a wide variety of current and past forces (Welner, 2001), so the greater involvement and effectiveness of public scholars would not be felt as a policy force in isolation. Those efforts will surely encounter obstacles and opposition—other forces shaping the policy context. Nonetheless, progress in this direction is a reasonable expectation, since the new research would become part of the conversation and, perhaps more importantly, part of a policy feedback loop. Currently, the feedback loop is dominated by outcome data such as graduation and college matriculation rates, in addition to an overabundance of test scores. These data feed into an achievement-gap conversation and achievement-gap policies, which then drive the need for even more outcome data to assess the impact of such policies.

That can change. Data concerning important inputs and processes can feed into an opportunity-gap conversation and opportunity-gap policies, and that conversation and those policies can drive the need for more input and process data. What is the path to that new routine? To create a different policymaking system with a research-based feedback loop—a system focused on closing

opportunity gaps—researchers must provide data and analyses about those gaps, giving measurement-focused policymakers greater capacity to address disparities in opportunity.

In different ways, the five challenges discussed above ask us to balance standards of research quality (e.g., preserving nuance) against very different standards of policy impact (e.g., generating and communicating actionable information). How should researchers negotiate these potentially conflicting standards and goals? We found no easy answers to this question. While it is common for research to involve methodological choices and compromises, researchers who deliberately engage with policy debates are asked to walk two or three additional steps down a methodologically uncomfortable road. In our own negotiation of these methodological challenges, we were guided by three principles: to be research-based, methodologically sound and transparent. These three principles, we argue, might offer insight to other projects that seek to influence education policy and practice.

Measuring Opportunity: Policy Relevant and Engaged Research

The last few decades have seen a variety of calls for researchers to become more publicly engaged (Nichols, 2007; Oakes, 2017), and we have seen the corresponding growth of diverse and creative approaches to scholarly engagement with communities, such as community-based and participatory research, action research, and design-based implementation research (Giles, 2008). In line with these efforts toward public scholarship, the OTLi aims to assemble actionable measures about the complex, multi-faceted dimensions of educational opportunity. In doing so, we are attempting to bridge interdisciplinary knowledge-building conversations with policy-making conversations. Public scholarship of this type demands the design and communication of research in ways that reach diverse audiences and that often address controversial policy issues. Here, our argument builds on older traditions of publicly engaged research, as well as a body of research that considers the distinctly political dimensions of how research frames policy debates (Coburn & Turner, 2012; Henig, 2012; McDonnell, 2009; Tseng, 2012; Weiss, 1978). An emerging literature on research utilization in policy also offers some insights for public scholarship. While often focused on how research is *used*—rather than *produced*—this field has explored the different discourse communities of research and policy (McLaughlin, 1987), how diverse actors across different sectors use research to drive educational practice and shape policy goals (Coburn & Turner, 2012; McDonnell, 2009), how policy makers acquire and interpret data (Asen & Gurke, 2014; Lubienski, DeBray, & Scott, 2014), and how research can be communicated more effectively to policymakers (Nutley et al., 2007). Scholars have also conceptualized research utilization as an issue of translation. As Tseng (2012) notes: "Because research does not speak for itself, policymakers and practitioners must always interpret its meaning and implications for their particular problems and circumstances. This means that identifying the right translators and creating productive conditions for translation are critical" (2012, p. 1).

Tseng and Nutley (2014) highlight how research can play a role in shaping different stages of the policy process, from the definition—and redefinition—of problems to the implementation of particular solutions. They echo an early insight from Carol Weiss (1977), who documented how research works, often in slow and unpredictable ways, to frame how policymakers think about problems and orient themselves to issues. This is, however, not a neutral process; rather, research use is irreducibly political (Henig, 2012). Any given policy decision is the result of many forces, with research sometimes included—but power, communications and politics often determine the research that makes it this far. In addition, while rarely conclusive, research is often employed as a powerful political weapon in policy debates (Scott et al., 2009). Yet rather than seeing politics as

something to be bracketed or eliminated from the research endeavor, we echo other scholars in asking researchers to strategically work "through politics rather than around it" (Henig, 2012, p. 13). The lines between research, advocacy of research, and advocacy of policy are undoubtedly tricky ones. But the research base needs thoughtful and persistent advocates who can frame issues in accessible and actionable terms.

A key element of such public scholarship calls upon researchers to join with others to more directly communicate research with policy-makers and practitioners, fostering democratic public problem solving in ways that resonate with broad concerns (Benson, Harkavy, & Puckett, 2007). In effect, public scholars are asked to engage multiple publics in multiple ways. Researchers, we contend, play a vital role in constituting and defining these publics around issues that merit attention. As Dewey (1927) argued nearly a century ago, science works when researchers engage in sustained, reciprocal public inquiry around concrete problems. The cross-disciplinary effort to connect opportunity-gap research with policy asks researchers to negotiate the challenges of communicating to wider audiences and the inescapably political nature of the knowledge enterprise. These are important challenges, especially in an increasingly partisan political environment. Here, a longstanding tradition of public scholarship may, if we choose to embrace it, provide a model for how rigorous research can engage with public policy.

References

American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*, *35*(6), 33–40. doi.org/10.3102/0013189X035006033

Asen, R., & Gurke, D. (2014). The Research on Education, Deliberation, and Decision-Making (REDD) Project. In K. Finnigan & A. Daly (Eds.). *Using Research Evidence in Education* (pp. 53–68). Springer International Publishing. doi.org/10.1007/978-3-319-04690-7_5

Au, W. (2010). The idiocy of policy: The anti-democratic curriculum of high-stakes testing. *Critical Education,* 1(1). http://m1.cust.educ.ubc.ca/journal/v1n1.

Au, W. (2011). Teaching under the new Taylorism: High-stakes testing and the standardization of the 21st century curriculum. *Journal of Curriculum Studies, 43*(1)*,* 25–45. doi.org/10.1080/00220272.2010.521261

Baker, B. D., Sciarra, D. G., & Farrie, D. (2015). *Is School Funding Fair? A National Report Card*. Education Law Center.

Ball, S. (1997). Policy sociology and critical social research: A personal review of recent education policy and policy research. *British Educational Research Journal, 23*(3)*,* 257–274. doi.org/10.1080/0141192970230302

Barnett, W. S. (2011). Effectiveness of early educational intervention. *Science, 333*(6045)*,* 975–978. doi.org/10.1126/science.1204534

Barnett, W. S., Friedman-Krauss, A. H., Weisenfeld, G. G., Horowitz, M., Kasmin, R. & Squires, J. H. (2017*). The state of preschool 2016: State preschool yearbook.* National Institute for Early Education Research. Retrieved from http://nieer.org/wp-content/uploads/2017/09/Full_State_of_Preschool_2016_9.15.17_compressed.pdf

Barnett, W. S., & Lamy, C. E. (2013). Achievement gaps start early: Preschool can help. In P. L. Carter & K. G. Welner (Eds.). *Closing the opportunity gap: What America must do to give every child an even chance* (pp. 98–110). New York, NY: Oxford University Press. doi.org/10.1093/acprof:oso/9780199982981.003.0007

Benson, L., Harkavy, I. R., & Puckett, J. L. (2007). *Dewey's dream: Universities and democracies in an age of education reform: Civil society, public schools, and democratic citizenship.* Philadelphia, PA: Temple University Press.

Berends, M., & Peñaloza, R. V. (2010). Increasing racial isolation and test score gaps in mathematics: A 30-year perspective. *Teachers College Record, 112*(4)*, 978–1007.

Berliner, D., (2009). *Poverty and potential: Out-of-school factors and school success.* Boulder, CO: Education and the Public Interest Center & Education Policy Research Unit. Retrieved from http://nepc.colorado.edu/publication/poverty-and-potential.

Blazar, D., & Pollard, C. (2017). Does Test Preparation Mean Low-Quality Instruction? *Educational Researcher, 46*(8)*, 420–433. doi.org/10.3102/0013189X17732753

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

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15*(5)*, 662–679. doi.org/10.1080/1369118X.2012.678878

Bush, G. W. (2004). *President's remarks in Minneapolis, Minnesota.* http://www.presidency.ucsb.edu/ws/?pid=72749

Camilli, G., & Firestone, W. A. (1999). Values and state ratings: An examination of the state-by-state education indicators in Quality Counts. *Educational Measurement: Issues and Practice, 18*(4)*, 17–25. doi.org/10.1111/j.1745-3992.1999.tb00278.x

Card, D., & Rothstein, J. (2006). *Racial segregation and the black-white test score gap.* Retrieved from the National Bureau of Economic Research website: http://www.nber.org/papers/w12078. doi.org/10.3386/w12078

Carlson, D., Borman, G. D., & Robinson, M. (2011). A multistate district-level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement. *Educational Evaluation and Policy Analysis, 33*(3)*, 378–398. doi.org/10.3102/0162373711412765

Carter, P. L. (2013). Student and school cultures and the opportunity gap: Paying attention to academic engagement and achievement. In P. L. Carter & K. G. Welner (Eds.). *Closing the opportunity gap: What America must do to give every child an even chance* (pp. 143–168). New York, NY: Oxford University Press. doi.org/10.1093/acprof:oso/9780199982981.003.0010

Carter, P. L., & Welner, K. G. (Eds.) (2013). *Closing the Opportunity Gap: What America Must Do to Give All Children an Even Chance.* New York, NY: Oxford University Press. doi.org/10.1093/acprof:oso/9780199982981.001.0001

Centers for Disease Control and Prevention. (n.d.). CDC's National Surveillance Data (1997-2015). Environmental Health Tracking Program and Lead Poisoning Prevention Program. https://www.cdc.gov/nceh/lead/data/index.htm.

The Child and Adolescent Health Measurement Initiative. (2013). Child Health Measures Content Map. U.S. Department of Health and Human Services. Retrieved from http://childhealthdata.org/docs/nsch-docs/2011-12-content-map-for-child-health-measures-pdf.pdf.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). *The academic achievement gap in grades three to eight.* Retrieved from the National Bureau of Economic Research website: doi.org/10.3386/w12207

Clotfelter, C. T. (1998). *Public school segregation in metropolitan areas.* Retrieved from the National Bureau of Economic Research website: doi.org/10.3386/w6779

Clune, W. H., & White, P. A. (2008). Policy Effectiveness of Interim Assessments in Providence Public Schools. WCER Working Paper No. 2008-10. *Wisconsin Center for Education Research (NJ1).*

Coburn, C. E. & Turner, E. O. (2012). Interventions to promote data use: An introduction. *Teachers College Record, 114*(11)*,* 1–13.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. (1966). *Equality of educational opportunity.* Washington, D.C.: Government Printing Office.

Coleman-Jensen, A., McFall, W., & Nord, M. (2013, May). *Food insecurity in households with children: Prevalence, severity, and household characteristics, 2010–11.* U.S. Department of Agriculture, Economic Research Service.

Collins, J. (2009). Social reproduction in classrooms and schools. *Annual Review of Anthropology, 38,* 33–48. doi.org/10.1146/annurev.anthro.37.081407.085242

Colorado Department of Education. (2017). District and School Performance Frameworks. Retrieved from https://www.cde.state.co.us/accountability/performanceframeworks.

Colorado Department of Education. (2015). School and District Performance Framework Overview. Retrieved from http://www.cde.state.co.us/accountability/spfdpf_technicalwriteup_072814.

Darling-Hammond, L. (1992). Educational indicators and enlightened policy. *Educational Policy*, *6*(3), 235–265. doi.org/10.1177/0895904892006003001

Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, *106*(6), 1047–1108. doi.org/10.1111/j.1467-9620.2004.00372.x

DeBray-Pelot, E., & McGuinn, P. (2009). The new politics of education analyzing the federal education policy landscape in the post-NCLB era. *Educational Policy*, *23*(1), 15–42. doi.org/10.1177/0895904808328524

Dewey, J. (1927). *The Public and Its Problems: An Essay in Political Inquiry.* New York, NY: Holt.

Du Bois, W. E. B. (1949). *The freedom to learn.* Lincoln University Press.

Elmore, R., & Fuhrman, S. (1995). Opportunity-to-learn standards and the state role in education. *The Teachers College Record*, *96*(3), 432–457.

Gamoran, A. (2001). American schooling and educational inequality: A forecast for the 21st century. *Sociology of Education*, *74,* 135–153. doi.org/10.2307/2673258

Giles, D.E., Jr., (2008). Understanding an emerging field of scholarship: Toward a research agenda for engaged, public scholarship. *Journal of Higher Education Outreach and Engagement, 12*(2), 97–106.

Guiton, G., & Oakes, J. (1995). Opportunity to learn and conceptions of educational equality. *Educational Evaluation and Policy Analysis*, *17*(3), 323–36. doi.org/10.3102/01623737017003323

Hazelkorn, E. (2015). *Rankings and the reshaping of higher education: The battle for world-class excellence.* New York, NY: Palgrave Macmillan. doi.org/10.1057/9781137446671

Henig, J. R. (2012). The politics of data use. *Teachers College Record, 114*(11)*,* 1–32.

Howe, K. R. (2003). *Closing methodological divides: Toward democratic educational research.* Boston, MA: Kluwer.

Howe, K. R. & Meens, D. E. (2012). *Democracy left behind: How recent education reforms undermine local school governance and democratic education.* Boulder, CO: National Education Policy Center. Retrieved from http://nepc.colorado.edu/publication/democracy-left-behind.

Howe, K. R. & Murray, K. (2015). *Why school report cards merit a failing grade.* Boulder, CO: National Education Policy Center. Retrieved from http://nepc.colorado.edu/publication/why-school-report-cards-fail.

Hoynes, H., Miller, D. & Simon, D. (2015). Income, the earned income tax credit, and infant health. *American Economic Journal: Economic Policy, 7*(1)*,* 172 –211. doi.org/10.1257/pol.20120179

Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. *Journal of public Economics, 89*(5-6)*,* 761–796. doi.org/10.1016/j.jpubeco.2004.08.004

Jin, G., & Whalley, A. (2007). *The power of information: Do rankings affect the public finance of higher education.* Retrieved from the National Bureau of Economic Research website: http://www.nber.org/papers/w12941.

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in US schools. *Educational Researcher, 35*(7)*,* 3–12. doi.org/10.3102/0013189X035007003

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. doi.org/10.3102/0034654308324427

Losen, D. J. (2003). Challenging racial disparities: The promise and pitfalls of the No Child Left Behind Act's race-conscious accountability. *Howard Law Journal*, *47,* 243–298.

Lubienski, C., DeBray, E., & Scott, J. (2014). The politics of research production, promotion, and utilization in educational policy. *Educational Policy, 28*(2)*,* 131–144. doi.org/10.1177/0895904813515329

Luca, M., & Smith, J. (2013). Salience in quality disclosure: Evidence from the U.S. News college rankings. *Journal of Economics & Management Strategy*, *22*(1), 58–77. doi.org/10.1111/jems.12003

Marr, C., Huang, C. C., Sherman, A., & DeBot, B. (2015). *EITC and Child Tax Credit promote work, reduce poverty, and support children's development, research finds.* Center on Budget and Policy Priorities. Retrieved from: https://www.cbpp.org/sites/default/files/atoms/files/6-26-12tax.pdf

McDermott, K. A. (2011). *High stakes reform: The politics of educational accountability.* Washington, DC: Georgetown University Press.

McDonnell, L. M. (2004). *Politics, persuasion, and educational testing.* Cambridge, MA: Harvard University Press.

McDonnell, L. M. (2009). Repositioning politics in education's circle of knowledge. *Educational Researcher, 38*(6)*,* 417–427. doi.org/10.3102/0013189X09342584

McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis, 9*(2)*,* 171–178. doi.org/10.3102/01623737009002171

McNall, M. A., Lichty, L. F., & Mavis, B. (2010). The impact of school-based health centers on the health outcomes of middle school and high school students. *American Journal of Public Health, 100*(9)*,* 1604–1610. doi.org/10.2105/AJPH.2009.183590

McNeil, L. (2000). *Contradictions of school reform: Educational costs of standardized testing.* New York, NY: Routledge.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook.* Thousand Oaks, CA: Sage.

Milner, H. R. (2013). Rethinking achievement gap talk in urban education. *Urban Education, 48*(1)*,* 3–8. doi.org/10.1177/0042085912470417

Murnane, R. J., & Duncan, G. J. (Eds.) (2011). *Whither opportunity? Rising inequality and the uncertain life chances of low-income children.* New York, NY: Russell Sage Foundation.

Nardo, M., Saisana, M., Saltelli, A., & Tarantola, S. (2005). *Tools for composite indicators building.* Ispra: European Commission..

Nichols, L. T. (Ed.). (2017). *Public sociology: The contemporary debate.* New York, NY: Routledge. doi.org/10.1017/9781316418369.033

Nutley, S., Walter, I., & Davies, H. T. O. (2007). *Using evidence: How research can inform public services.* Bristol, UK: Policy Press. doi.org/10.2307/j.ctt9qgwt1

Oakes, J. (1989). What educational indicators? The case of assessing the school context. *Educational Evaluation and Policy Analysis, 11*(2), 182. doi.org/10.3102/01623737011002181

Oakes, J. (2005). *Keeping track: How schools structure inequality* (2nd ed.). New Haven, CT: Yale University Press.

Oakes, J. (2017). 2016 AERA Presidential Address: Public Scholarship: Education Research for a Diverse Democracy. *Educational Researcher.* doi.org/10.3102/0013189X17746402

Orfield, G., & Frankenberg, E. (2014). Increasingly segregated and unequal schools as courts reverse policy. *Educational Administration Quarterly*, *50*(5), 718–734. doi.org/10.1177/0013161X14548942

Orr, A. (2003). Black-white differences in achievement: The importance of wealth. *Sociology of Education, 76*(4), 281–304. doi.org/10.2307/1519867

Patton, M. Q. (2008). *Utilization-focused evaluation.* Los Angeles, CA: Sage.

Piketty, T. (2013). *Capital in the 21st century.* Cambridge, MA: Belknap Press.

Planty, M. (2010). *Understanding education indicators: A practical primer for research and policy.* New York, NY: Teachers College Press.

Propper, C., & Wilson, D. (2003). The use and usefulness of performance measures in the public sector. *Oxford Review of Economic Policy, 19*(2), 250–267. doi.org/10.1093/oxrep/19.2.250

Quint, J., Sepanik, S., & Smith, J. (2008). Using student data to improve teaching and learning: Findings from an evaluation of the Formative Assessments of Students Thinking in Reading (FAST-R) program in Boston elementary schools. New York, NY: MDRC.

Reardon, S. F. (2011). The widening socioeconomic status achievement gap: New evidence and possible explanations. In R. J. Murnane & G. J. Duncan (Eds.), *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children.* New York, NY: Russell Sage Foundation.

Scott, J., Lubienski, C., & DeBray-Pelot, E. (2009). The politics of advocacy in education. *Educational Policy*, *23*(1), 3–14. doi.org/10.1177/0895904808328530

Scott, J. T., & Wells, A. S. (2013). A more perfect union: Reconciling school choice policy with equality of opportunity goals. In P. L. Carter & K. G. Welner (Eds.). *Closing the opportunity gap: What America must do to give every child an even chance* (pp. 123–140). New York, NY: Oxford University Press. doi.org/10.1093/acprof:oso/9780199982981.003.0009

Shepard, L. A. (1991). Negative policies for dealing with diversity: When does assessment and diagnosis turn into sorting and segregation. *Literacy for a diverse society: Perspectives, practices, and policies*, 279–298.

Simon, H. A. (1957). *Models of man; social and rational.* New York, NY: Wiley.

Slavin, R. E., Cheung, A., Holmes, G., Madden, N. A., & Chamberlain, A. (2013). Effects of a data-driven district reform model on state assessment outcomes. *American Educational Research Journal, 50*(2), 371–396. doi.org/10.3102/0002831212466909

Thompson, J. (2011). Did Data-Driven Accountability Cause the Atlanta Cheating Scandal? *The Huffington Post.* Retrieved March 5, 2013, from http://www.huffingtonpost.com/john-thompson/did-datadriven-accountabi_b_894179.html.

Tseng, V. (2012). The uses of research in policy and practice. *Social Policy Report*, *26*(2), 3–23.

Tseng, V., & Nutley, S. (2014). Building the infrastructure to improve the use and usefulness of research in education. In Finnigan, K. S., & Daly, A. J. (Eds.). *Using research evidence in education* (pp. 163–175). Springer International Publishing. doi.org/10.1007/978-3-319-04690-7_11

U.S. Department of Health and Human Services. (2007). *Toxicological profile for lead*. Public Health Service Agency for Toxic Substances and Disease Registry: Atlanta, GA. Retrieved from http://www.atsdr.cdc.gov/toxprofiles/tp13.pdf.

Verbeeten, F. H. (2008). Performance management practices in public sector organizations: Impact on performance. *Accounting, Auditing & Accountability Journal, 21*(3), 427–454. doi.org/10.1108/09513570810863996

Weiss, C. H. (1977). Research for policy's sake: The enlightenment function of social research. *Policy Analysis, 3,* 531–545.

Weiss, C. H. (1978). Improving the linkage between social research and public policy. In L. E. Lynn (Ed.) *Knowledge and policy: The uncertain connection* (pp. 23–81). Washington, DC: National Academy of Sciences.

Welner, K. G. (2001). *Legal rights, local wrongs: When community control collides with educational equity.* Albany, NY: SUNY Press.

Welner, K. G. (2015, June 10). "The business of rankings: Did the *US News & World Report* make substantial mistakes?" *The Conversation.* Online at https://theconversation.com/the-business-of-rankings-did-the-us-news-and-world-report-make-substantial-mistakes-42678.

Welner, K. G., & Mathis, W. J. (2015). *Reauthorization of the Elementary and Secondary Education Act: Time to Move beyond Test-Focused Policies.* Boulder, CO: National Education Policy Center.

Wiley, T. G., & Wright, W. E. (2004). Against the undertow: Language-minority education policy and politics in the "age of accountability". *Educational Policy, 18*(1), 142–168. doi.org/10.1177/0895904803260030

Willinsky, J. (2001). The strategic education research program and the public value of research. *Educational Researcher, 30*(1), 5–14.

Wright, W. E. (2002). The effects of high stakes testing in an inner-city elementary school: The curriculum, the teachers, and the English language learners. *Current Issues in Education, 5*(5). doi.org/10.3102/0013189X030001005

About the Authors

Ryan H. Pfleger
Newark Public Schools
rhpfleger@gmail.com
Ryan Pfleger is a researcher at Newark Public Schools, New Jersey, where he began as a Strategic Data Project Fellow with the Center for Education Policy Research at Harvard University. His research explores the connections between economic, racial, and educational inequality.

Terri S. Wilson
University of Colorado Boulder
Terri.Wilson@colorado.edu
Terri S. Wilson is an assistant professor in the School of Education at the University of Colorado Boulder. Her research focuses on the philosophical foundations of education policy, including issues raised by school choice, marketization, and parent engagement.

Kevin G. Welner
University of Colorado Boulder
Kevin.Welner@colorado.edu
Kevin Welner is a professor in the School of Education at the University of Colorado Boulder, specializing in educational policy. He is director of the National Education Policy Center, housed at CU Boulder.

Charles Bibilos
cbibilos@gmail.com
Charles Bibilos has nearly 20 years of experience as a teacher, curriculum developer, and education policy consultant. He holds an M.A. in Education from the University of Colorado Boulder.

education policy analysis archives

arquivos analíticos de políticas educativas
conselho editorial