

The Landscape of School Rating Systems

Ben Dalton



RTI Press publication OP-0046-1709

This PDF document was made available from www.rti.org as a public service of RTI International. More information about RTI Press can be found at <http://www.rti.org/rtipress>.

RTI International is an independent, nonprofit research organization dedicated to improving the human condition by turning knowledge into practice. The RTI Press mission is to disseminate information about RTI research, analytic tools, and technical expertise to a national and international audience. RTI Press publications are peer-reviewed by at least two independent substantive experts and one or more Press editors.

Suggested Citation

Dalton, B. (2017). *The Landscape of School Rating Systems*. RTI Press Publication No. OP-0046-1709. Research Triangle Park, NC: RTI Press. <https://doi.org/10.3768/rtipress.2017.op.0046.1709>

This publication is part of the RTI Press Research Report series. Occasional Papers are scholarly essays on policy, methods, or other topics relevant to RTI areas of research or technical focus.

RTI International
3040 East Cornwallis Road
PO Box 12194
Research Triangle Park, NC
27709-2194 USA

Tel: +1.919.541.6000
E-mail: rtipress@rti.org
Website: www.rti.org

©2017 RTI International. All rights reserved. Credit must be provided to the author and source of the publication when the content is quoted. No part of this publication may be reproduced in any form by any electronic or mechanical means without permission in writing from the publisher. RTI International is a registered trademark and a trade name of Research Triangle Institute.

<https://doi.org/10.3768/rtipress.2017.op.0046.1709>

www.rti.org/rtipress

Contents

About the Author	i
Abstract	ii
Introduction	1
School Ratings Systems in the Context of Comparative Organizational Assessments	2
Impacts of School Ratings, Accountability, and Performance Metrics on School Practices	3
A Review of State and Consumer-Oriented School Rating Systems	4
State Ratings Systems	5
Consumer-Oriented Rating Systems	9
Issues in the Design of School Ratings Systems	11
Measurement	11
Transformation	12
Integration	13
Presentation	14
Conclusion and Implications	16
References	17

About the Author

Ben Dalton, PhD, is a Senior Education Research Analyst at RTI International. Dr. Dalton is project director of *US News & World Report's* Best High Schools Rankings project and Co-Principal Investigator of *The Role of Industry-Recognized Credentials in High School Completion and Postsecondary Enrollment in the State of Florida*, an Institute of Education Sciences (IES) research grant.

Abstract

The rise of the accountability movement in education has resulted in the proliferation of school report cards, school ratings and rankings, and other kinds of performance reporting for public consumption and policy use. To understand the strengths and limitations of school rating systems and the role they play in shaping public perceptions and school improvement practices, this paper situates rating systems within the broader field of comparative organizational assessments and neo-institutional theory; describes school rankings and rating systems in use by states and consumer-oriented enterprises; and details four aspects of school ratings (measurement, transformation, integration, and presentation) that affect their use and interpretation.

Introduction

The rise of the accountability movement in education has resulted in the proliferation of school report cards, school ratings and rankings, and other kinds of performance reporting for public consumption and policy use (Coburn & Turner, 2012). These performance reports emerge from a long history of increasing quantification of the performance of private enterprises and public agencies (Espeland & Sauder, 2007) and are intended to increase transparency and lower asymmetries of information between the public and seemingly opaque educational agencies (Gormley & Weimer, 1999; Haertel & Herman, 2005). School performance reports became widespread due to the passage of No Child Left Behind (NCLB), the 2001 re-authorization of the federal Elementary and Secondary Education Act (ESEA), which mandated public reporting of standardized test score performance for public-sector schools. With the waiver of strict NCLB reporting requirements initiated in 2011 and the re-authorization of ESEA as the Every Student Succeeds Act (ESSA) in 2015, accountability systems and school ratings have diversified further, drawing on increasingly detailed data systems and the dissemination of advanced statistical techniques to merge test performance and other kinds of measures into more comprehensive school assessments (ESSA, 2015; US Department of Education, 2015).

These reporting systems have engendered a host of positive and negative consequences for school and teacher practices, public administration, and family choice (Booher-Jennings, 2005; Colyvas, 2012; Diamond & Cooper, 2007; Hastings & Weinstein, 2008). School administrators and individual teachers have adapted to the pressure of high-stakes accountability reporting in various ways, from systematic attempts to improve instruction to manipulating test results or other reports of school performance (Heilig & Darling-Hammond, 2008; Herman & Haertel, 2005). In part to address the unintended consequences of performance reporting, accountability policies have shifted away from a narrow set of increasingly punitive responses to poor school performance to more comprehensive and ongoing supports to all schools to improve the

education they provide (Martin, Sargrad, & Batel, 2016).

To understand the role that rating systems play in guiding accountability supports and policy decisions, and the effects that performance ratings and rankings have on teacher and administrator practices within schools, it is important to understand what the ratings or rankings are intended to accomplish and the key methodological and design decisions that are involved in crafting them. To contribute to this understanding, this paper situates school rating systems within the broader field of comparative organizational assessments and neo-institutional theory; describe school rankings and rating systems in use by states and consumer-oriented enterprises; and detail four aspects of school ratings systems that affect their use and interpretation. Examining school rating systems comparatively and in light of broader work on organizational assessments can provide an opportunity to foster deeper, more meaningful conversation about the appropriate uses of performance measures for school and student improvement.

Paying particular attention to high schools, we focus on school ratings systems that provide a score, grade, rank, or other rating to individual schools based on their performance on various student outcome measures. This definition excludes accountability or school improvement categories such as NCLB's designations related to failing to meet adequate yearly progress (AYP) multiple years in a row, and similar classifications implemented by states in response to approval of NCLB waivers and currently being designed and implemented under ESSA. These designations are part of larger accountability systems that tie performance measurement (including, oftentimes, summary school ratings), accountability categories, and school supports or interventions together. A full rendering of the landscape of accountability systems is beyond the scope of this paper, but the usefulness and efficacy of school ratings must be ultimately evaluated in this fuller context. That said, school ratings themselves have significant implications for schools and educational agencies. To understand why this is the case, it is helpful to begin by delineating how school ratings fit

within the wider landscape of performance reports known as comparative organizational assessments.

School Ratings Systems in the Context of Comparative Organizational Assessments

Elementary and secondary school ratings and ranking systems are one example of a much broader phenomenon of comparative organizational assessments that have grown dramatically in recent decades in response to the explosion in available data and the increasing ease of computing and publishing (frequently online) ever more complex evaluations of organizations, institutions, and government bodies (Coe & Brunet, 2006). Comparative organizational assessments are any type of cross-organization or cross-institution report of comparable metrics on performance, processes, rules, resources, or other factors relevant to the evaluation of the target enterprise (Gormley & Weimer, 1999). The target institution can either be (a) governing jurisdictions such as countries, states, counties, school districts, or other agencies at the same level; or (b) individual organizations such as hospitals, other health service providers, early child care providers, elementary or secondary schools, colleges or universities, graduate departments, businesses or corporations, or charities. Assessments are produced across many fields of public policy and governance, including health care, economic policy, the environment, and education (see Coe & Brunet [2006] for a review of some prominent reports, for example).

In addition to applying to different kinds of organizations and fields, comparative assessments can take many different forms, characterized by both the technical quality of the measures involved and more qualitative aspects of their design that improve communicability and heighten impact (Gormley & Weimer, 1999; Stinchcombe, 2001). These forms include organizational report cards, “scorecards,” benchmarking, rankings, ratings, or some combination of these (Coburn & Turner, 2012; Gormley & Weimer, 1999; Kaplan & Miyake, 2010; Matthews, 1998). Reports may be published as a single compendium document covering all assessed organizations (particularly when the number of organizations is small, as when states are given

“report cards”) or as individual documents per organization—both types of which are increasingly made available online in fixed or interactive formats.

A key distinction among different types of comparative organizational assessments is between those which provide summative ratings (a score, grade, or rank) and those which provide multiple comparative metrics without an overall rating. Scorecards (including the “balanced scorecard” approach of Kaplan and Miyake [2010]) and many states’ school report cards initially created in response to the requirements of NCLB fall in the latter category. For example, California’s School Accountability Report Cards (SARCs) and California School Dashboard provide information about school enrollment, demographics, student performance on state tests, annual yearly progress (AYP) determinations, graduation rates, Title I status, school staff data, expenditures, and a variety of other information, including facilities information and physical fitness test results—without providing any single grade or rating of the school’s performance across measures. In contrast, summative rating systems yield an easily understood overall assessment of the organization’s quality or performance. Because of the reduction of an organization’s processes and outcomes to a single rating or grade, summative ratings are highly visible outcomes that can exert a strong influence on public perceptions, organizational goals, and individual and corporate practice (Jacobsen, Snyder, & Saultz, 2014). Although report cards or scorecards that do not report a summary measure can be useful for internal management and organizational improvement and may be better avenues for strategic planning and administrative purposes, their potential to have a galvanizing impact on consumer behavior and public policy is much lower (Gormley & Weimer, 1999).

School ratings are thus an individual, summative type of comparative organizational assessment applied to public and private schools at all levels. These rating systems use various combinations of student performance, student population, organizational resources, and other factors to create an overall summary grade, score, ranking, or rating that explicitly evaluates individual schools or districts

relative to others within a jurisdiction or group of comparable organizations. The oldest, most well-known and well-studied of such systems are *US News & World Report's* (*US News*) annual Best Colleges rankings, which use surveys of college and university administrators and publicly available information from the federal government to rank institutes of higher education within peer groups (such as National Universities and Regional Colleges). A variety of other higher education rankings and ratings (such as rankings of graduate programs or schools) have also been produced by other organizations and by *US News* itself. At the elementary and secondary levels, there are multiple ratings of public schools produced by states and organizations such as Greatschools.org, *US News*, *Newsweek*, and Niche.com, in addition to state-level report cards such as *Education Week's* Quality Counts and the federal government's own ranking of state performance on the National Assessment of Educational Progress (self-labeled as, in fact, "the Nation's Report Card").

Comparative organizational assessments serve different purposes depending on the reporting organization and the nature of the assessment. The "balanced scorecard" approach, for example, is designed not only as a mechanism for tracking progress and ensuring accountability but is also intended to align with strategic goals of specific organizations (Kaplan & Miyake, 2010; Muller, 2015); indeed, they are primarily for internal use, not for external evaluation and comparison. Coe and Brunet (2006) argue that different types of report issuers (governments, commercial enterprises, academics, foundations, and public interest groups) strongly shape the goals and design of assessments. Public interest groups in the environmental field, for example, sometimes pursue a strategy of dramatizing failure to heighten alarm about a particular issue and drive legislative and policy agendas. Although such manipulation is not prevalent in the field of education, it is important to consider the different purposes of organizations such as Greatschools.org (a nonprofit that licenses its ratings to other organizations, such as real estate website Zillow.com), *US News* (a journalistic publication and website), and individual states (which possess different resources,

challenges, and policy priorities) in interpreting their ratings.

Taken together, the target field and institution, the technical and qualitative design features of the assessments—specifically whether a summative rating is provided—and the issuer's goals and resources all influence the nature of the comparative assessment and its usefulness to consumers, policymakers, and organizational leaders. With this in mind, it is helpful to turn toward some of what research has uncovered about the genesis of and reaction to comparative organizational assessments in education.

Impacts of School Ratings, Accountability, and Performance Metrics on School Practices

In addition to the literature describing the ideal uses and construction of organizational assessments (Gormley & Weimer, 1999; Moynihan, 2008), a burgeoning research literature has sought to understand the impact of accountability reporting on educators and educational institutions. In general, this research draws from theoretical perspectives in organizational theory, particularly the new institutionalism of sociology, which emphasizes how such formal codes and reported measures (everything from national flags to economic statistics, and including educational assessments) promote uniformity in outward appearance (isomorphism) but coexist with local, contextualized practices that are resistant to external pressure (DiMaggio & Powell, 1983; Meyer & Rowan, 1977). In other words, institutionalism as applied in organizational theory has observed a decoupling between the formal structures organizations claim adherence to and the actual routines and beliefs of the organization and its employees. This allows organizations to gain legitimacy in the eyes of the public or principal constituencies as members in good standing while minimizing disruption to preferred modes of business. According to this perspective, the development of organizational metrics, report cards, and ratings would yield superficial conformity to overall institutional goals (as expressed in the chosen

measures), but not necessarily change attitudes or behaviors within local contexts.

However, research into the impact of comparative organizational assessments within education challenges this narrative. For example, Espeland and Sauder (2007; Sauder & Espeland, 2009) found that law school administrators felt both external and internal, self-imposed pressure to align their organizations with the measures promoted by *US News's* rankings of best law programs. Such effects can be positive or negative; numerous studies have found that the emphasis on accountability and testing codified by NCLB within elementary and secondary schools has led to negative effects such as ignoring low-achieving students, reducing arts and other kinds of enrichment instruction, encouraging low-achievers to skip the tests, and even criminal cheating practices (Booher-Jennings, 2005; Colyvas, 2012; Diamond & Cooper, 2007; Heilig & Darling-Hammond, 2008). Both “gaming” ratings systems and significant changes in teaching emphasis have been observed; the former could have been predicted by institutionalism, while the latter indicates “recoupling,” or the alignment between individual organizational procedures and external institutional forms.

Likewise, investigations into the use of data within schools has suggested that educators react in a variety of ways to accountability policies and performance reporting. Teachers can experience both cultural and technical barriers to data use that limit the impact of performance reporting on actual practice (Ingram, Louis, & Schroeder, 2004) as well as come to adopt the conceptual framework of performance-based accountability in their thinking about and approach to classroom instruction and student development (Spillane, Parise, & Sherer, 2011). More ominously, in adapting themselves to the pressures of external accountability, schools have been found to “game the system” of accountability to, for example, exclude likely low-performing students (Heilig & Darling-Hammond, 2008) or even change test answers and scores of students (Wilson, Bowers, & Hyde, 2011). The important point is that organizational assessments, and school performance reports in particular, seem to exhibit “tight coupling” with the actual activities of practitioners. Both resistance to

and embrace of data tools and accountability thinking can lead to shifts in organizational routines (Spillane, 2012).

This conclusion implies that rating systems must be understood in terms that go beyond the statistical or technical aspects of their construction. Qualities outlined by Gormley and Weimer (1999) and Stinchcombe (2001)—such as validity, comprehensibility, and the intent and capacity of issuing organizations—play a role in the reception of ratings and the reaction of practitioners. We must apply conscientious attention to the construction, use, and misuse of ratings systems if they are to have a constructive impact in school improvement. To further this goal, the paper next turns to a closer examination of state and consumer-oriented high school rating systems and then the key technical qualities that affect their use and interpretation.

A Review of State and Consumer-Oriented School Rating Systems

Multiple organizations now release school ratings. There are two primary types of issuers: (1) state departments of education or public instruction, which report school ratings relative to other schools within their state (Martin et al., 2016); and (2) consumer-oriented enterprises which publish ratings or rankings for schools across the country. Consumer-oriented enterprises include the nonprofit Greatschools.org (which is supported by advertising and licensing revenue as well as foundations and grants) and journalistic organizations including *US News*, *The Washington Post*, and *Newsweek*, and commercial websites such as Niche.com. Also, a variety of award programs have produced assessments of schools or districts that meet some of the criteria for a rating system, except a summative rating. Such efforts include the Department of Education's Blue Ribbon Schools Program awards, which awards distinctions to public and private high schools based on academic performance and gap reduction; the Broad Prize for Public Education and the Broad Prize for Charter Management Organizations, which until recently presented awards to public school districts or charter organizations, respectively, demonstrating

high overall performance and reduced achievement gaps; and the State Collaborative on Reforming Education (SCORE) Prize, which recognizes high-performing schools and districts in Tennessee. These awards rely on a mix of quantitative analysis and qualitative investigation but do not necessarily report publicly on their internal analyses.

State Ratings Systems

All states publish school report cards on their public elementary and secondary schools; however, not all states have implemented a rating or grade that summarizes the school's overall performance (not including accountability categories or formerly required NCLB designations of "meeting" or "not meeting" AYP). States will be required to add these summative ratings under ESSA. Indeed, several states are currently revisiting their accountability systems in light of ESSA requirements and regulations.

Although they predate NCLB in some states, most states' school report cards, scorecards, and performance reports arose in response to NCLB requirements that school performance be publicly reported. In 2011, recognizing the flaws in NCLB's requirements that all students be proficient by 2014, the US Department of Education initiated a waiver program that allowed states freedom from some NCLB requirements in exchange for a plan to implement new systems of college and career readiness assessment, accountability, teacher and principal evaluations, and low-burden administrative reporting (US Department of Education, 2015). Many states took this opportunity to revamp or expand their school ratings systems. As of the fall of 2015, 44 states¹ had requested and received approval for NCLB flexibility, with another two requests (Iowa and Wyoming) under review at the time. Only California, Montana, Nebraska, North Dakota, and Vermont did not request waivers.

With the passage of ESSA in December 2015, federal requirements for accountability changed. States are now required to report information beyond

assessment results; disaggregate data by additional student subgroups; place "substantial weight" on certain indicators; and report a summative rating. The indicators ESSA requires include test performance in English language arts (ELA) and mathematics; a second academic indicator such as achievement growth; progress toward achieving English language proficiency; one or more school quality (e.g., school climate) or student success measures (i.e., college and career readiness measures); and, for high schools, graduation rates. These requirements are to take effect for the 2017–2018 school year.

Based on a review of state education websites, accountability documents, school-level report cards, and NCLB waiver requests, Table 1 presents information (as of June 2017) on the use of summative school ratings in each of the states, the components of the summary ratings, and the name the state has given the core rating. The components listed are given generic names that cover a variety of specific measures that states may use. For example, "achievement" can include multiple subject areas (although usually just English and mathematics); "progress" refers to both student-level achievement growth models and year-over-year increases in proficiency among aggregate student groups; and "college and career readiness" includes a variety of measures such as performance on Advanced Placement (AP) or International Baccalaureate (IB) courses, SAT or ACT test scores, and matriculation rates at postsecondary institutions. The listing of components includes elements specific to high schools, such as graduation rates and college and career readiness indicators; elementary and middle school ratings in most states are identical except for the inclusion of these metrics.

Fifteen states provide no summative ratings for their schools or do not have published plans for a summative rating. These states do report a variety of profile and outcome measures at the school level—and in some cases, like Texas, report summative ratings for certain dimensions like achievement or growth—but they have not taken the additional step of constructing an overall summary rating for schools. In other cases, individual districts or

¹ Count includes Washington, DC, but not Puerto Rico or the Bureau of Indian Education.

Table 1. Summary of school rating systems in use or planned by states

State	Summary rating	Components of summary rating	Name
Alabama	A–F	achievement, growth, gaps, graduation, attendance, and college and career readiness	School Grades
Alaska	Stars (1–5)	achievement, growth, gaps, graduation, attendance, and college and career readiness	Alaska School Performance Index
Arizona	A–F	achievement, growth, graduation, and dropout	A–F Letter Grade Accountability System
Arkansas	A–F	achievement, growth, gaps, graduation	Arkansas School Grading System
California	None		
Colorado	0%–100% with classification	achievement, growth, gaps, graduation, dropout, and postsecondary and workforce readiness	School Performance Framework
Connecticut	0–100 with classification	achievement, growth, graduation, college and career readiness, absenteeism, physical education, arts access	Next Generation Accountability System
Delaware	0–500	achievement, growth, graduation, and college and career readiness	Delaware School Success Framework
District of Columbia	0–100+ with classification	achievement, growth, graduation, attendance, and test participation	School Index Score
Florida	A–F	achievement, growth, college and career readiness, and graduation rate	School Grades
Georgia	A–F	achievement, growth, gaps, and bonus measures	College and Career Readiness Performance Index (CCRPI)
Hawaii ¹	0–400 with classification	achievement, growth, gaps, and college and career readiness	Strive HI School Performance Report
Idaho	None		
Illinois	0–300 with classification (forthcoming)	achievement, growth, progress, graduation, bonus measures, and test participation	Multiple Measures Index
Indiana	A–F	achievement, growth, graduation, and college and career readiness	PL221 (Public Law 221) Grades
Iowa	0–100 with classification	achievement, growth, gaps, graduation, and staff retention	Performance Index
Kansas	None		
Kentucky ²	Classification based on matrix (forthcoming)	achievement, gap, graduation, college and career readiness, and opportunity to learn	To be determined
Louisiana	A–F	achievement, growth graduation, college readiness, and bonus measures	School Performance Score
Maine	A–F	achievement, growth, and graduation	Maine School Performance Grading System
Maryland ¹	Strands (1–5) based on 0–2+ scale	achievement, growth, gaps, and college readiness	School Progress Index
Massachusetts	Levels (5–1) with labels	achievement, growth, gaps, graduation, dropout, college and career readiness, and bonus measures	Progress and Performance Index
Michigan	Colors (5)	achievement, progress, gaps, and graduation rate	Top-to-Bottom ranking
Minnesota ²	0%–100%	achievement, growth, gaps, and graduation	Multiple Measures Rating
Mississippi	A–F	achievement, growth, graduation, and college and career readiness	Accountability Grades

(continued)

Table 1. Summary of school rating systems in use or planned by states (continued)

State	Summary rating	Components of summary rating	Name
Missouri	0–100	achievement, graduation, college and career readiness, and attendance	Annual Performance Report (APR) Score
Montana	None		
Nebraska	None		
Nevada ¹	Stars (1–5)	achievement, growth, gaps, graduation, college and career readiness, and attendance	Nevada School Performance Framework
New Hampshire	None		
New Jersey	None		
New Mexico	A–F	achievement, growth, graduation, college and career readiness, opportunity to learn, and bonus measures	School Grading
New York	None		
North Carolina	A–F	achievement and growth	School Performance Grades
North Dakota	None		
Ohio	A–F (forthcoming)	achievement, growth, gaps, graduation, and college and career readiness	Ohio School Report Cards
Oklahoma	A–F	achievement, growth, and bonus measures of graduation, college and career readiness, and attendance	A–F School Grading System
Oregon	None		
Pennsylvania	0–100+ with colored icons	achievement, growth, gaps, graduation, attendance, and bonus measures of college and career readiness	Building Level Academic Score
Rhode Island	20–100 with classification	achievement, gaps, progress, and graduation	Composite Index Score
South Carolina ^{1,3}	A–F	achievement, progress, graduation, and test participation	
South Dakota	0–100 with classification	achievement, growth, graduation, and college and career readiness	School Performance Index
Tennessee	None		
Texas	None		
Utah	A–F	achievement, growth, graduation, college and career readiness, and test participation	School Grade
Vermont	None		
Virginia	None		
Washington	Levels (1–10)	achievement, growth, graduation, and college and career readiness	Washington Achievement Index
West Virginia	A–F	achievement, growth, graduation, college and career readiness, and attendance	West Virginia School Accountability System
Wisconsin	0–100 with classification	achievement, growth, gaps, graduation, college and career readiness, attendance, and test participation	Overall Accountability Score
Wyoming	None		

1 For school year 2015–2016 (the most recent year of reporting), Hawaii, Maryland, Nevada, Oregon, and South Carolina did not implement their rating system, pending revisions in response to ESSA.

2 Minnesota planned to suspend their school rating system in July 2017, pending revisions in response to ESSA. Kentucky continued reporting their old ratings (a 0–100 index with classification) in July 2017, but plans a transition to a new system.

3 South Carolina suspended its ratings after the 2015–2016 school year.

Note: Components include measures that only apply to high schools (graduation; college and career readiness and its variants).

Sources: Elementary and Secondary Education Act (ESEA) Flexibility Waivers; State school report cards and accountability websites; Martin, Sargrad, & Batel (2016).

regional educational agencies may report their own grading system: although not systematically covered here, one example is the California Office to Reform Education's (CORE) consortium of districts in northern California, which applied for and received its own NCLB waiver to implement a "School Quality Improvement Index" that uniquely incorporates measures of social-emotional skills as outcomes. Similar other educational groups that are specific to states may report their own summative rating: for example, the Connecticut Coalition for Achievement Now (ConnCAN) assigns A through F letter grades to Connecticut schools and its web pages are prominent in search results, even though Connecticut's state system provides a 0–100 index rating.

The remaining 36 states formally provide summative ratings or have plans to do so for all of its schools. The ratings the states provide can be divided into three groups:

- **A–F letter grades.** The most common rating system is a letter grade rating familiar to parents and consumers. Florida was the first state to implement a letter grading system; currently, 16 states use letter grading systems or have plans to do so. A–F grading systems typically have underlying continuous values or scales to which letter grades are subsequently assigned. The underlying values may or may not be clearly explained in accountability documentation or presented along with the letter grade in public documentation.
- **Index scores or scales.** Thirteen states use continuous scores or scales to summarize school performance. Most of these (10 states) range from 0 or 1 to 100 (or slightly more, depending on whether bonus points are awarded to schools for performance on additional measures). Three states use larger ranges of 0–300 (Illinois), 0–400 (Hawaii), or 0–500 (Delaware). States that use index scores also typically have a classification system that translates the continuous score into a category that is used for improvement purposes. For example, Washington, DC, reports five categories based on its 0–100+ School Index Scores: Reward, Rising, Developing, Focus, and Priority. Other states use more distinctive labels. Pennsylvania, for example,

has a unique system of colored icons associated with ranges of its 0–100+ scale (e.g., an upward-facing, hollow blue triangle is used with the highest range of scores).

- **Other ratings.** Seven states use alternative summary ratings. Two states (Alaska and Nevada) use a five-star rating system. Three states use numerically defined levels: Maryland calls its five levels "strands," based on an underlying z-score that ranges from 0 to approximately 2. Massachusetts uses levels 5 to 1, with level 1 being the highest, and color-codes each level from green (level 1) to red (level 5). Washington uses levels 1 (low) to 10 (high). Kentucky uses a classification scheme based on a matrix of categories and resulting in categories such as "Intervention" and "Outstanding." Michigan uses a five-color scheme as its summative rating (green, lime, yellow, orange, and red, with green as the highest level).

These summative ratings may or may not be prominent on school report cards, or may be reported on web sites that are separate from other accountability indicators. For example, before suspending their summative rating for 2015–2016, South Carolina did not include their letter grade rating on school report cards (an effort to revamp South Carolina's reporting system is outlined in Koon, Petscher, & Hughes, 2015).

Despite differences in the labels attached to the ratings, the similarities across states are striking. Most states using a summative rating are using a five-level rating system, whether that is designated by stars, levels, or classes based on a continuous metric. In addition, the component indicators used to rate schools are very similar. By following NCLB flexibility request requirements and with the technical assistance of the US Department of Education, states prepared a variety of similar accountability plans that typically incorporated at least several of the following measures: overall achievement within in a school (the unadjusted or absolute level of performance), achievement growth among the same students over time, achievement progress of students or the school toward defined goals, achievement gaps between

key student subgroups, graduation rates, attendance rates, and college and career readiness results. Under ESSA, states have more flexibility to design their accountability systems, including incorporating additional measures beyond these core indicators. For example, Kentucky proposes to include opportunity measures relating to equitable access to gifted and talented programs and certified teachers (Kentucky Department of Education, 2017).

Schools usually receive points for their performance on each measure, with their total score comprising a summation across measures. The score then either may total the final index (as in Hawaii) or be converted to a standard range or percentage (as in Colorado). The total possible points received for any given measure typically varies; some measures are given more weight in the total score by having more possible points awarded for performance in that domain. Some measures have multiple indicators, and receive points for each indicator contributing to the total points for that measure. For example, the measure of achievement usually includes separate components for reading and mathematics test performance.

As a detailed example, Nevada's School Performance Framework (currently on hold pending revisions in response to ESSA) created a five-star rating system based on a 100-point scale involving achievement status (accounting for 20 percent of the scale and comprising the percentage of 10th graders meeting proficiency expectations and the cumulative percentage of 11th graders meeting proficiency expectations); school median growth percentile for 10th graders (10 percent); gap in 11th-grade proficiency for disadvantaged student groups (10 percent) (each of the foregoing using both math and reading assessments); graduation rate (30 percent, including both an overall rate and an indicator for the gap in rates for disadvantaged student groups); college and career readiness (16 percent, using four different measures involving Advanced Placement proficiency, ACT and SAT test participation, advanced diploma rates, and Nevada college remediation rates); and two other measures (average daily attendance [10 percent]

and percentage of ninth graders who are credit deficient [4 percent]). Nevada's star rating system was thoroughly explained on the state Department of Education's website (<http://nspf.doe.nv.gov/>) and was featured prominently on individual school pages, but it was not featured on a separate set of school report card pages also operated by the Nevada Department of Education (<http://www.nevadareportcard.com>).

Consumer-Oriented Rating Systems

As noted, consumer-oriented ratings are produced by the nonprofit organization Greatschools.org as well as by journalistic publications and websites such as US News and Niche.com. Like state rating systems, Greatschools.org, Niche.com, and SchoolDigger.com rate elementary and middle schools as well as high schools; but *US News*, *Newsweek*, and *The Washington Post* only cover high schools. Unlike state rating systems, these organizations rate schools across states, and therefore must grapple with state tests that are not comparable across state lines and proficiency benchmarks that may be set at different levels across states. The typical solution is to provide some rating or measure that accounts for school performance relative to the school's state average. They may also incorporate comparable cross-state indicators such as the percent of graduating seniors who took or passed Advanced Placement courses. They are not able to take advantage of student-level data available in states' unit-record longitudinal databases; instead, they rely on aggregate information at the school level.

The Greatschools Rating

The largest and most prominent independent school rating system is the nonprofit Greatschools.org, which reports its ratings on its own website as well as licenses its rating to other organizations (for example, it is featured on real estate website Zillow.com). Greatschools.org reports a 1–10 rating and labels scores 1–3 as below average, 4–7 as average, and 8–10 as above average. In most states, ratings are based only on test performance relative to other schools in the state. Schools are rated on how well their students (in the aggregate) do on each test reported; then the individual test ratings are averaged into the

overall 1–10 rating. In 12 states, Greatschools.org has additional data on academic growth and college readiness (which they define as graduation rates and SAT or ACT performance). Schools in these states are also compared relative to other schools in the state, and then averaged with test performance into an overall rating—if only one additional indicator is available, then test performance is weighted at half the overall average, with the other indicator as the other half. If both additional indicators are available, then the three components are equally weighted as one-third of the total.

US News & World Report's Best High Schools Rankings

US News ranks a subset of high schools that meet certain criteria (US News & World Report, 2017). The ranking is based on an overall rating—a “college readiness index”—calculated from participation and passing rates on Advanced Placement (AP) and International Baccalaureate (IB) tests; this criterion-referenced measure allows cross-state comparisons and a ranking of the best high schools across the country and within individual states. However, only high schools that perform better than expected given their poverty level (using a regression analysis and predicted proficiency results) and better than their state’s average for disadvantaged subgroups are ranked on the college readiness index. This “gated” methodology excludes approximately two-thirds of high schools, consistent with the “best” focus of the rankings. For those that are ranked, US News also awards gold, silver, and bronze “medals” to further distinguish classes of high schools.

Newsweek's America's Top High Schools

Newsweek produces two ranked lists of top high schools in the United States: an “absolute” list based on unadjusted performance on state tests (relative to other schools within each state) and a “relative” list that takes into account the percentage of low-income students (Newsweek, 2015). As the first step in constructing each list, Newsweek also uses a “gated” method. For the absolute list, Newsweek uses a “threshold” analysis to identify the top 20 percent of high schools in each state. For the relative list, it identifies those schools that perform half of a standard deviation above the state average after

controlling for poverty. Once these schools are identified, Newsweek uses the results of its own high school survey to construct a college readiness score based on enrollment rate (accounting for 25 percent of the total score); graduation rate (20 percent); weighted AP/IB composite (17.5 percent); weighted SAT/ACT composite (17.5 percent); holding power, which is the change in enrollment between 9th and 12th grades (10 percent—this last component comes from the Common Core of Data, not Newsweek’s survey); and counselor-to-student ratio (10 percent). In this regard, their ranking is similar to state composite indices. However, schools without sufficient data from the survey, or who are nonrespondents, are not included. In 2015—the most recent year for which a methodology report was released—Newsweek reported that their survey response rate was 34 percent for schools on the relative list and 42 percent for schools on the absolute list.

Other Rankings

Given the widespread availability of data available from the states and from the federal government, there is no shortage of school ratings or rankings available to the public. Others worth mentioning include *The Washington Post's* Most Challenging High Schools index, which ranks schools on a “Challenge Index,” the ratio between the number of AP, IB, and Advanced International Certificate in Education (ICE) tests taken in a given school year and the number of seniors who graduate in May or June (Matthews, 1998). Niche.com also offers high school ratings, and incorporates multiple measures drawn from federal data as well data from Niche users (students, alumni, and parents) who provide feedback at their website. Their letter-grade rating (A–D) incorporates measures of academics, health and safety, student culture and diversity, teachers, resources and facilities, clubs and activities, sports and fitness, and opinion items from survey responses. Each measure has its own subcomponents and variable weighting. The underlying scale is a z-score created from the z-scores of each individual measure. SchoolDigger.com offers ratings of elementary and secondary schools based on normalizing and averaging schools’ state-reported test performance, similar to Greatschools.org.

Issues in the Design of School Ratings Systems

The overview of state and consumer-oriented ratings systems presented above demonstrates a wide variety of methodologies and modes of presentation for high school ratings, despite underlying similarities in the reliance on a core set of metrics related to student test scores and a few other immediate “output” measures like graduation rates. The variety of ways in which ratings are presented reinforces the idea that presentation is as consequential as measures and methods. Indeed, in their work *Organizational Report Cards*, Gormley and Weimer (1999) argue that there are six key aspects of comparative assessments: validity, comprehensiveness, relevance, comprehensibility, reasonableness, and functionality. Tellingly, only the first feature relates to the technical qualities of the measures themselves; the remaining five features relate to how ratings systems effectively summarize and communicate useful information to or about the organizations being assessed. Likewise, Colyvas (2012) has criticized prior work concerning what she terms performance metrics for focusing too closely on the statistical properties of measures. Colyvas draws on Stinchcombe’s (2001) framework for analyzing formal systems to articulate features (several similar to Gormley and Weimer’s list) that have a significant bearing on the meaning and use of such measures. Reyna (2016) also recently discussed key features of school ratings systems in the environment of ESSA.

To further understanding of the importance of nontechnical features in the design of school ratings systems, the following discussion elaborates on the broader measurement, methodological, and presentation choices made in constructing them. Understanding these choices can help advance knowledge of the limitations and strengths of school ratings in guiding school improvement and informing parents and the public about the schools their children attend. Drawing on and consolidating the aspects discussed at length by Gormley and Weimer (1999), Colyvas (2012), Stinchcombe (2001), and Reyna (2016), the discussion focuses on the implications of four design features of school rating

systems: (1) measurement, (2) transformation, (3) integration, and (4) presentation.

Measurement

The movement toward educational accountability has mainly stressed outcomes measures: primarily achievement on standardized assessments and, at the high school level, attainment (Martin et al., 2016; Mikulecky & Christie, 2014). However, there are multiple types of data on school performance that could be measured and which could provide useful information to practitioners and families. Gormley and Weimer (1999), for example, describe inputs, processes, outputs, and outcomes as elements of organizational performance that can be measured and reported on. Their “outputs” are what are often labeled “outcomes” in the educational field—immediate or short-term results such as test performance. “Outcomes,” in their terminology, refer to longer-term results that reflect the ultimate mission of an organization, such as successfully enabling a transition to college and providing the basis for good work and family habits. Although “college and career readiness” outcomes are often included in state rating systems (Martin et al., 2016), they are typically only measured through coursetaking or college admissions test results that are, in fact, immediate or short-term results, not explicit evidence of college or work success (Mikulecky & Christie, 2014).

In addition to outputs and outcomes, inputs are also an important aspect of organizational performance, reflecting organizational competencies in securing resources to meet needs. These aspects of school performance are not regularly reported on school report cards nor included in accountability systems, but may be reported independently in financial and budgetary documents. There is certainly far less regular, systematic reporting of school-level organizational resources (e.g., number, age, and diversity of textbooks; amount, quality, and appropriateness of high school laboratory equipment), and very little attention has been paid to how to achieve consistency in reporting on and informing the public about school resources.

Likewise, process measures have received very scant attention within school ratings systems, yet are clearly the critical component in what makes schools work for students and their families. Being among the most difficult aspects of organizations to measure, and whose complexity is most in danger of being reduced to unfairly simplistic metrics, it is understandable that measures of process—such as teacher professional development, instructional improvement plans, and discipline procedures—are rarely incorporated into school rating systems. An exception is Kentucky’s “Unbridled Learning” system, which incorporates program reviews and professional ratings into their accountability system, if not into their school ratings.

To be fair, accountability policies never intended to encompass inputs and processes—these were in many ways assumed to adjust to meet accountability goals (Hoffer, 2000)—but the expansion of accountability reporting into full-scale school rating systems incorporating many measures, and tied more strongly to purposeful systems of intervention and improvement, implies that school leaders and policymakers need access to higher-quality data on more than just short-term outputs or even long-term outcomes. Standardized school-level summaries of instructional materials and technology; facilities; teacher and leadership evaluations; course assignment and access policies; and discipline and safety procedures and practices—among many other possible options—deserve their place in rating systems that purport to provide an overview of school quality and performance.

Transformation

The selection of domains to measure and specific indicators to use leads to a second step involving the transformation of data to make measures amenable for analysis and ready for inclusion in a summary score or rating. Depending on the issuer of a school rating, measures may be included that reflect absolute or unadjusted performance (for example, overall achievement results) or conditioned measures that account for the “inputs” involved in their production (such as predicted achievement levels, academic growth measures at the individual level, or change in academic achievement at the school or student subgroup levels).

Indeed, Gormley and Weimer place a strong emphasis on the necessity of making appropriate “risk adjustments”—that is, controlling for differences in populations served and prior histories. Such statistical adjustments are critical to ensuring the validity of organizational performance metrics, which, according to Gormley and Weimer, should measure the contributions of the organization itself and not the educational or socioeconomic backgrounds of students. This includes, in the school context, removing the influence of peers and the composition of the study body, long shown to have an effect on individual student achievement (Wilkinson, 2002; Sacerdote, 2014). Statistical controls can be relevant for cross-school comparisons by addressing factors outside of the school’s control that are known to affect student performance, or for same-school comparisons over time by accounting for prior performance level and changing student composition across cohorts.

However, while Gormley and Weimer note that validity must be balanced by comprehensibility, their emphasis on modeling of educational outcomes to abstract out organizational performance ignores the function that peer groups or student composition might play in some interpretations of school ratings. That is, risk adjustment may be seen by some ratings consumers as an unnecessary confusion or distraction. Unadjusted performance could be very meaningful and valid to parents who believe in the power of peers and the importance of school climate for fostering good educational habits. Even those who speak out from concern about subgroup performance may decide that adjustments hide performance problems and result in setting low expectations for underperforming groups (Gormley & Weimer, 1999, p. 77).

Fortunately, most school ratings systems combine some measure of unadjusted achievement with an adjusted measure of growth or progress to both provide credit to schools for their performance as an organization given resource constraints or population deficits and to acknowledge the public’s concern with overall performance and high standards. Indeed, the inclusion of performance gaps in most school rating systems indicates that the tide has shifted away from earlier performance reporting thinking (as illustrated

by Gormley & Weimer, 1999) that schools should be judged relative to the challenges they face toward a system in which schools are rewarded for across-the-board (absolute) levels of achievement.

The other principal data transformation that has an important consequence on school ratings results is whether the data are preserved as a criterion-referenced measure (e.g., a graduation rate measured as 0 to 100) or are standardized into relative metrics of percentiles, z-scores, or relative ranks. This dimension of relativity is treated differently in state school rating systems and consumer-oriented rating systems. In the former, schools are typically given credit for the extent to which they meet external benchmarks—e.g., percent proficient on math and reading assessments, or graduation rate—and all schools can theoretically attain the highest possible rating. In the latter, consumer-oriented ratings, including Greatschools.org and US News rankings, schools are given a relative rating, inherently limiting the number of schools that can receive top ratings. In Greatschools.org's case, schools are rated relative to others in the state on their final letter grade rating. In the *US News* rankings, schools are given a final ranking nationally (and by state) but are also given a criterion-referenced score on its "college readiness index," on which all schools can (mathematically) achieve a top score.

The different purposes of state and consumer-oriented ratings partially account for differences in the use of relative metrics. State ratings are designed for accountability and improvement purposes; consumer publications are designed to clearly differentiate schools to help guide relocation or school choice decisions and drive media attention. However, reward or incentive programs (such as South Carolina's school incentive reward program; Gormley & Weimer, 1999, pp. 77–80) conducted by states, districts, or independent agencies (such as the Broad Foundation) also often use relative metrics as a way to create distinctions for the purposes of identifying single winners or clear dividing lines for groups of awardees.

Integration

Another significant design feature of school rating systems is the manner in which multiple measures are jointly used or integrated. At the broadest level is the question of whether multiple measures are integrated within a single unidimensional rating or if measures are kept separate as part of a multidimensional evaluation of quality and performance. Multidimensional rating systems include approaches such as the balanced scorecard or any of the currently 14 state accountability systems that provide no unitary score or grade. Multidimensional rating systems are better suited for internal purposes of evaluation and improvement, as they demand greater knowledge of individual measures or domains of measurement; they are less accessible to outside observers such as parents or state or federal policymakers (see Gormley and Weimer's discussion of "comprehensibility").

Even if the school rating system is holistic in the sense that it incorporates measures across multiple domains and employs some element of statistical adjustment for inputs, if the final result is a score, grade, or rank, the rating is a unidimensional evaluation metric. Among unidimensional ratings, there is the aforementioned distinction between gated evaluations (e.g., *US News's* Best High Schools rankings) and composite scoring (used by most states). Gated evaluations are ideal for systems whose purpose is to identify top-performing schools for the purposes of distinction or reward. Composite scoring serves the need for comprehensive performance evaluations of all schools for accountability and improvement.

There also remains the question of the construction of composite scoring systems, which can follow several different approaches. The construction most commonly employed by state rating systems and some consumer-oriented publications (Greatschools.org and Niche.com) are weighted schemes, in which each measure contributes a varying proportion of a final index or score. Weighting has strong advantages because it allows multiple measures to contribute to a rating (in contrast to the "all or nothing" results of NCLB's AYP provisions) while emphasizing measures deemed most significant, but

there are few standard mechanisms for setting relative weights other than qualitative processes of expert review, public comment, and political input (Reyna, 2016). For example, ESSA stipulates only that states are required to give “substantial weight” to a set of academic indicators in designing their accountability systems, and that these academic indicators receive more weight than school quality or student success measures (Martin et al., 2016). Indeed, the lack of clear procedures for determining weights means that some ratings systems are subject to ongoing debate that may be influenced by political or other noneducational concerns (e.g., as in North Carolina—see Antoszyk, 2017).

There are also systems in which measures could contribute to placement in a categorical level or rating based on decision rules tied to multiple measures instead of a single continuous score or index. The special case here is Texas’s Performance Index system, which although not providing an overall continuous summary score, computes index values for four domains (achievement, progress, gaps, and postsecondary readiness) and then assigns schools a “met standard” or “improvement required” classification based on surpassing target scores on each dimension. An alternative method could combine weighted composites with categorical label adjustment by adjusting a categorical rating determined by a continuous score upwards or downwards depending on additional measures, such as level of test participation (Reyna, 2016).

Presentation

Presentation is one of the key aspects of rating systems and performance metrics that both Stinchcombe (2001) and Gormley and Weimer (1999) stress. As an aspect of comprehensibility (in Gormley and Weimer’s term) or communicability (in Stinchcombe’s term), presentation encompasses both the naming or labeling of the rating and any associated categories as well as the manner in which it is disseminated. Prior research has shown that parents and the public respond to the presentation of performance data and the ways that performance data are summarized. Hastings and Weinstein (2008) showed that parents were more likely to select higher-

performing schools when given performance data. Jacobsen and colleagues (2014, p. 17) found that letter grades have a strong influence on public perceptions of school performance:

For the strong school, the letter grade format leads to strongly positive views of school performance, while other formats lead to positive but more mediocre views of school performance. While we may have expected the numerical formats to be more influential due to cultural notions about numbers being accurate and precise...our consistent and strong results for the letter grade condition suggest that other cultural measures can have a much stronger influence. It appears people believe an A indicates very high performance and a C indicates quite low performance, while the numbers are more ambiguous.

Similarly, school rankings—which, unlike ratings such as A–F grades, explicitly sort schools into higher or lower positions—may suggest larger differences between schools than the underlying measure actually indicates. *US News*, for example, uses a series of tiebreakers to account for the fact that many of the highest-ranked schools earn the maximum college readiness index score of 100. Further, the actual distinction between schools separated by the same number of positions (i.e., rank 10 and rank 20, versus rank 100 and rank 110) may be variable. Although ratings are not immune to misinterpretation as to the actual underlying performance of schools—indeed, scores, grades, or other categories of performance have an implicit hierarchy—rankings reify performance distinctions that may not be meaningful or clear-cut.

Likewise, the manner of promotion or dissemination is signally important; ratings that are not visible or easily accessible will receive scant attention. For example, some states do not currently include available ratings on their school report cards (such as Rhode Island’s Composite Index); or they may provide ratings through separate websites or reports (such as Hawaii’s rating, which is only presented on its more detailed “School Status and Improvement Reports”). Even within states that include summary ratings, their prominence can differ sharply—Oklahoma’s and New Mexico’s letter grades, for example, are featured in large font on the top of the

first page of their school report cards. The chosen presentation can deemphasize or promote the school rating and profoundly affect how students, parents, educators, and the public think about their local schools and the public school system overall.

Table 2 presents a summary of some of these key features for the major high-profile high school rating systems. For comparison, the required ESSA accountability system, the historical NCLB system, and the prize funds rating systems (e.g., the Broad Foundation and the Tennessee SCORE Prize) are also included. NCLB instituted a system of school-adjusted accountability designations based on meeting benchmarks, primarily on a single dimension of test performance (but also including graduation rate at the high school level; because of the emphasis on test scores and to distinguish NCLB more clearly from subsequent state accountability systems under granted NCLB waivers, NCLB is labeled a unidimensional system). The prize funds also do not typically create a single composite rating or score, but rank schools or districts internally on a variety of metrics to identify potential award winners and vet finalists.

These design features describe methodological choices that must be made in creating ratings for schools. Of course, these choices are not the only ones, nor necessarily the most consequential ones, involved in designing school ratings or scores. For example, the first and foremost process is generating individual measures themselves, whether drawn from existing databases or created during a survey or other process of data collection. These processes are not unbiased or error-free, and any adequate rating system must consider the role and impact of known or suspected problems with the validity and reliability of measures to be included. A second and vitally important set of choices informs how the rating system feeds back into decision-makers' and educators' actions—the “reactive” process discussed by Espeland and Sauder (2007). Depending on the purpose of the ratings, the issuer, the nature of its design, and how ratings are used by stakeholders, systems must take into account how practices may be amplified or distorted through the highlighting effect of the ratings and their measures. These and other processes are not design choices, but can be equally or more important factors in the development, use, and effectiveness of school ratings systems.

Table 2. Design features of school ratings systems

System/issuer	Type of rating	Adjustments	Relativity	Dimensionality	Rating methodology
NCLB requirements	Accountability category	School-level	Criterion-referenced	Unidimensional	Benchmarking
States with NCLB waivers					
Without a summary rating	Accountability category	Student-level	Criterion-referenced	Multidimensional	Benchmarking
With a summary rating	Composite rating plus accountability category	Student-level	Criterion-referenced	Multidimensional	Variable weighting
ESSA requirements	Composite score plus accountability category	Student-level	Criterion- and/or norm-referenced	Multidimensional	Variable weighting
Greatschools.org ¹	Composite score	None	Norm-referenced	Unidimensional	Summary scale
Niche.com	Composite score	None	Norm-referenced	Multidimensional	Variable weighting
<i>US News & World Report</i>	Ranking plus index	School-level	Norm-referenced	Multidimensional	Gated
<i>Newsweek</i>	Ranking plus index	School-level	Norm-referenced	Multidimensional	Gated
Prize funds ²	Award	School-level	Norm-referenced	Multidimensional	Gated

Notes: ESSA = Every Student Succeeds Act; NCLB = No Child Left Behind.

¹ Refers to main Greatschools.org rating, not ratings based on additional measures for certain states that have provided Greatschools data beyond achievement data.

² Includes Tennessee SCORE Prize and Broad Foundation Prizes.

Conclusion and Implications

This article has reviewed state and consumer-oriented rating systems, situating them in the broader context of comparative organizational assessments and the research literature surrounding their use and impact. It has also discussed key design features and their implications for use and interpretation by educators, families, and policymakers.

Although school ratings do not necessarily inform parents about the quality of the instruction a child is likely to receive from individual teachers and although ratings are not pure representations of organizational performance given schools' student populations and resource constraints, they can provide useful information for guiding improvement and helping families make schooling choices. School rating systems are influential with school and district administrators, who pay close attention to public perception, and with parents, who use them as a guide for residential location and enrollment decisions. Because of school rankings' visibility and influence on policymaking and practice, their content, design, and presentation are important for educators and users to understand.

However, more research needs to be conducted on the construction and validity of school rating systems. As relatively new measures in the organizational assessments landscape, school ratings—particularly those supplied by states—do not have a large literature evaluating whether their ratings or rankings correlate with desired student outcomes or have unintended effects (see the literature on college rankings, e.g., Bastedo & Bowman, 2009; McDonough, Lising, Walpole, & Perez, 1998; Meredith, 2004). Although accountability systems in general (particularly NCLB) have been studied extensively (e.g., de Wolf & Janssens, 2007; Dee, Jacob, & Schwarz, 2013), there is a need for additional research into the effects of summative ratings and rankings on student performance, school and teacher practices, and perceptions of parents.

A national dialogue about appropriate local, state and federal accountability systems should draw on comparative analyses of different systems to identify promising improvements and remedy

potential flaws. The shift away from a single federally designed accountability system to a laboratory of varying rating systems among the states provides an opportunity to analyze the processes associated with constructing, maintaining, and improving school-level accountability systems. Combined with the rapid growth in education data and student-level longitudinal databases within states, this shift suggests the need for a new wave of studies about the relationship between educational governance and school performance—studies that may be able to show more conclusive results than those conducted in the nascent stages of national accountability (e.g., Hanushek & Raymond, 2005).

In addition, the overlap between prominent consumer-oriented ratings like GreatSchools.org and state rating systems deserves further attention, given the potential for public confusion caused by the availability of multiple ratings and the likelihood that some groups of users are more likely to receive their impressions of school performance and quality from some sources than from others. The role that various circulating ratings play in shaping the attitudes of different socioeconomic or racial/ethnic groups of parents requires further investigation. Such investigation would benefit those who manage educational systems, and those who are publicly committed to supporting and strengthening schools.

Finally, in creating ratings that improve schools and bolster public education, both public agencies and private enterprises have a role to play. Gormley and Weimer (1999), in the conclusion of their work on organizational report cards, note that the public and private sectors each have strengths to contribute to the dissemination of high-quality information to the public. Public agencies can marshal the resources and will to gather and validate the data necessary for institutional evaluations, while private, media-oriented businesses have expertise in creating compelling presentations and promoting wide dissemination of results. Although their aims differ, and the goals of different levels of educational policy and governance may differ as well, each has a compelling interest in providing meaningful, actionable data that can lead to an improved quality of life for children and their parents.

References

- Antoszyk, E. (2017). School report cards. *Education NC*. Retrieved from <https://www.ednc.org/map/2015/06/school-report-cards/>
- Bastedo, M. N., & Bowman, N. A. (2009). US News & World Report college rankings: Modeling institutional effects on organizational reputation. *American Journal of Education*, 116(2), 163–183.
- 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>
- Coburn, C., & Turner, E. O. (2012). The practice of data use: An introduction. *American Journal of Education*, 118(2), 99–111. <https://doi.org/10.1086/663272>
- Coe, C. K., & Brunet, J. R. (2006). Organizational report cards: Significant impact or much ado about nothing? *Public Administration Review*, 66(1), 90–100. <https://doi.org/10.1111/j.1540-6210.2006.00558.x>
- Colyvas, J. A. (2012). Performance metrics as formal structures and through the lens of social mechanisms: When do they work and how do they influence? *American Journal of Education*, 118(2), 167–197. <https://doi.org/10.1086/663270>
- Dee, T. S., Jacob, B., & Schwartz, N. L. (2013). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis*, 35(2), 252–279. <https://doi.org/10.3102/0162373712467080>
- de Wolf, I. F., & Janssens, F. J. (2007). Effects and side effects of inspections and accountability in education: An overview of empirical studies. *Oxford Review of Education*, 33(3), 379–396.
- Diamond, J. B., & Cooper, K. (2007). The uses of testing data in urban elementary schools: Some lessons from Chicago. In P. A. Moss (Ed.), *Evidence and decision making* (pp. 241–263). Malden, MA: National Society for the Study of Education. <https://doi.org/10.1111/j.1744-7984.2007.00104.x>
- DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2), 147–160. <https://doi.org/10.2307/2095101>
- Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, 113(1), 1–40. <https://doi.org/10.1086/517897>
- Every Student Succeeds Act (ESSA). (2015). Pub. L. No. 114–95.
- Gormley, W. T., Jr., & Weimer, D. L. (1999). *Organizational report cards*. Cambridge, MA: Harvard University Press.
- Haertel, E. H., & Herman, J. L. (2005). A historical perspective on validity arguments for accountability testing. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data for educational accountability and improvement: 104th yearbook of the National Society for the Study of Education, pt. 2* (pp. 1–34). Malden, MA: Blackwell. <https://doi.org/10.1111/j.1744-7984.2005.00023.x>
- 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>
- Hastings, J. S., & Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly Journal of Economics*, 123(4), 1373–1414. <https://doi.org/10.1162/qjec.2008.123.4.1373>
- Heilig, J. V., & 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>
- Herman, J. L., & Haertel, E. H. (Eds.). (2005). *Uses and misuses of data for educational accountability and improvement: 104th yearbook of the National Society for the Study of Education, pt. 2*. Malden, MA: Blackwell.
- Hoffer, T. B. (2000). Accountability in education. In M. T. Hallinan (Ed.), *Handbook of the sociology of education* (pp. 529–544). New York, NY: Kluwer.
- Ingram, D., Louis, K. R. S., & Schroeder, R. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record*, 106(6), 1258–1287. <https://doi.org/10.1111/j.1467-9620.2004.00379.x>

- Jacobsen, R., Snyder, J. W., & Saultz, A. (2014). Informing or shaping opinion? The influence of school accountability data format on public perceptions of school quality. *American Journal of Education*, 121(1), 1–27. <https://doi.org/10.1086/678136>
- Kaplan, R. S., & Miyake, D. N. (2010). The balanced scorecard. *School Administrator*, 67(2), 10–21. Retrieved from <http://www.aasa.org/SchoolAdministratorArticle.aspx?id=11784>
- Kentucky Department of Education. (2017). *Kentucky Accountability System (proposed) overview*. Retrieved from https://education.ky.gov/comm/Documents/III_AccountabilitySummaryDRAFT013117FINAL_0%20%281%29.pdf
- Koon, S., Petscher, Y., & Hughes, J. (2015). *Development and examination of an alternative school performance index in South Carolina (REL 2015–097)*. Washington, DC: US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from <https://ies.ed.gov/ncee/edlabs>
- Martin, C., Sargrad, S., & Batel, S. (2016). *Making the grade: A 50-state analysis of school accountability systems*. Washington, DC: Center for American Progress. Retrieved from <https://cdn.americanprogress.org/wp-content/uploads/2016/05/17094420/AccountabilityLandscape-report2.pdf>
- Matthews, J. (1998). *Class struggle: What's wrong (and right) with America's best public high schools*. New York, NY: Times Books.
- McDonough, P. M., Lising, A., Walpole, A. M., & Perez, L. X. (1998). *College rankings: Democratized college knowledge for whom? Research in Higher Education*, 39(5), 513–537. <https://doi.org/10.1023/A:1018797521946>
- Meredith, M. (2004). Why do universities compete in the ratings game? An empirical analysis of the effects of the *US News and World Report* college rankings. *Research in Higher Education*, 45(5), 443–461. <https://doi.org/10.1023/B:RIHE.0000032324.46716.f4>
- Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83(2), 340–363. <https://doi.org/10.1086/226550>
- Mikulecky, M., & Christie, K. (2014). *Rating states, grading schools: What parents and experts say states should consider to make school accountability systems meaningful*. Denver, CO: Education Commission of the States. Retrieved from <http://www.ecs.org/docs/rating-states,grading-schools.pdf>
- Moynihan, D. (2008). *The dynamics of performance management: Constructing information and reform*. Washington, DC: Georgetown University Press.
- Muller, R. J. (2015, June). Balanced scorecard: A tool for monitoring progress. *School Administrator*. Retrieved from <http://www.aasa.org/content.aspx?id=37342>
- Newsweek*. (2015). Identification of *Newsweek's* 2015 top 500 public high school rankings. Retrieved from <http://www.newsweek.com/methodology-newsweeks-top-high-school-rankings-2015-363698>
- No Child Left Behind Act of 2001 (NCLB). (2001). Pub. L. No. 107–110, 115 Stat. 1460.
- Reyna, R. (2016). *Key issues in aggregating indicators for accountability determinations under ESSA*. Washington, DC: Council of Chief State School Officers. Retrieved from <http://www.ccsso.org/Documents/2016/ESSA/KeyIssuesinAggregatingIndicators.pdf>
- Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: Two steps forward? *Annual Review of Economics*, 6(1), 253–272. <https://doi.org/10.1146/annurev-economics-071813-104217>
- Sauder, M., & Espeland, W. N. (2009). The discipline of rankings: Tight coupling and organizational change. *American Sociological Review*, 74(1), 63–82. <https://doi.org/10.1177/000312240907400104>
- Spillane, J. P. (2012). Data in practice: Conceptualizing the data-based decision-making phenomena. *American Journal of Education*, 118(2), 113–141. <https://doi.org/10.1086/663283>
- Spillane, J. P., Parise, L. M., & Sherer, J. Z. (2011). Organizational routines as coupling mechanisms: Policy, school administration, and the technical core. *American Educational Research Journal*, 48(3), 586–619. <https://doi.org/10.3102/0002831210385102>
- Stinchcombe, A. L. (2001). *When formality works: Authority and abstraction in law and organizations*. Chicago, IL: University of Chicago Press.
- US Department of Education. (2015). *ESEA flexibility page*. Retrieved from <https://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html>

- US News & World Report*. (2017). Identifying top-performing high schools for the “best high schools” rankings: Analytical methodology and technical appendices. Retrieved from <https://www.usnews.com/pubfiles/best-high-schools-technical-appendix.pdf>
- Wilkinson, I. A. G. (2002). Peer influences on learning: Where are they? *International Journal of Educational Research*, 37(5), 395–401. [https://doi.org/10.1016/S0883-0355\(03\)00012-0](https://doi.org/10.1016/S0883-0355(03)00012-0)
- Wilson, R. E., Bowers, M. J., & Hyde, R. L. (2011). *Special investigation into test tampering in Atlanta’s school system*. Atlanta, GA: Georgia Office of the Governor. Retrieved from https://archive.org/stream/215252-special-investigation-into-test-tampering-in/215252-special-investigation-into-test-tampering-in_djvu.txt

RTI International is an independent, nonprofit research institute dedicated to improving the human condition. Clients rely on us to answer questions that demand an objective and multidisciplinary approach—one that integrates expertise across the social and laboratory sciences, engineering, and international development. We believe in the promise of science, and we are inspired every day to deliver on that promise for the good of people, communities, and businesses around the world.

The RTI Press complements traditional publication outlets by providing another way for RTI researchers to disseminate the knowledge they generate. This PDF document is offered as a public service of RTI International.