Westat

Improving lives through research™

Identification of *Newsweek*'s 2016 Top 500 Public High School Rankings

Authors

Matthew Finster
Shen Lee

2016

Identification of *Newsweek's* 2016 Top 500 Public High School Rankings

2016

Prepared for:
Newsweek LLC
7 Hanover Square, 5th Floor
New York, New York 10004

Prepared by:
Westat
An Employee-Owned Research Corporation®
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

Executive Summary

This report describes the 2016 methodology and analysis for *Newsweek*'s "Top Public High School" rankings. The 2016 methodology is similar to *Newsweek's* 2014 and 2015 rankings in that it produced two sets of rankings: an absolute list and a relative list. For 2016, we again created two lists to demonstrate the consequence of accounting for non-school factors on school rankings. The absolute list ranks schools solely on the achievement and college readiness indicators. The relative list ranks the highest performing schools after accounting for student poverty levels. These two lists reveal how the rankings vary when non-school factors are considered.

For both the absolute and relative rankings, we conducted multiple analyses to identify the "top" schools. Using data from National Center for Education Statistics (NCES), specifically ED*Facts* and Common Core of Data (CCD), we conducted a threshold analysis to identify schools with the highest levels of academic achievement based on average student proficiency as measured by scores on state standardized assessments. The absolute list identified the top 30 percent of schools in each state, that is, schools with performance above the 70th percentile. The relative list identified schools that were .5 standard deviation (SD) or higher than the line of best fit after controlling for the percentage of students eligible for free or reduced-price lunch. We surveyed schools identified in the threshold analysis to collect college readiness data. The web-based survey asked for basic demographic information, graduation rates, college enrollment rates, number of full-time equivalent (FTE) counselors, number of students taking the SAT and ACT, average SAT and ACT scores, percentage of students taking at least one advanced placement (AP) course, percentage of students taking one international baccalaureate (IB) course, percentage of students taking one advanced international certificate of education (AICE) course, the schools' students' AP/IB/AICE scores, number of students participating in dual enrollment courses, and number of dual enrollment credits earned. Using these data, we developed a college readiness index score based on six variables: counselor FTE (weighted at 10 percent), changes in 9th-grade to 12th-grade enrollment rates (10 percent), a composite SAT/ACT score (17.5 percent), a composite AP/IB/dual enrollment score (17.5 percent), high school graduation rates (20 percent), and college enrollment rates (25 percent). The rankings are based on schools' college readiness index scores. For the absolute list, schools are ranked according to their college readiness index score. For the relative list, schools are ranked according to their college readiness index score, controlling for student poverty.

.

Contents

Exhibit

Figure

Introduction

The purpose of this brief is to present the methodology (hereafter termed 2016 methodology) used to develop *Newsweek*'s 2016 High School Rankings. The 2016 methodology remains largely unchanged from the 2014 and 2015 methodology. Similar to 2014 and 2015, the methodology produced rankings based on two aspects of school performance: the academic achievement of all students on state assessments in reading and mathematics and the extent to which high schools prepare their students for college. This year we, again, developed two rankings, an absolute and relative ranking. One of the biggest challenges in ranking school performance is addressing whether, and how, to incorporate the substantial influence of family background on students' success in school. Numerous studies have confirmed that family background has a considerable influence on student achievement and concluded that students from more affluent families experience substantial advantages over students from less affluent families (Sirin, 2005; Stinebrickner & Stinebrickner, 2003; Sutton & Soderstrom, 1999). Thus, schooling outcomes are influenced by a combination of factors inside and outside of school. Generally, high school rankings ignore the influence of family characteristics.[1] If family background characteristics are not accounted for, all student performance is erroneously attributed to the school. As a result, removing the influence of students' socioeconomic status on school-level outcomes can drastically change the high school rankings (e.g., see Toutkoushian & Curtis, 2005). To address the common criticism that school rankings are driven by student background characteristics, we developed two lists: one referred to as the "absolute" list and another as the "relative" list. The absolute ranking is based on the highest scores on the college readiness index and ranks schools without accounting for student background characteristics. The relative ranking, on the other hand, accounts for students' socioeconomic status to clarify the association between influences occurring outside of school and student performance. Persons interested in knowing which schools perform the best in readying students for college, regardless of whether the performance is attributable to the school or to family background, should refer to the absolute list. On the other hand, those interested in knowing which schools perform the best on the college readiness index relative to their levels of student socioeconomic status should refer to the relative list.

For 2016, we again developed an equity measure for the schools on both lists to assess economically disadvantaged students' performance. Long-standing conceptions of equity and adequacy in education finance (e.g., Baker & Green, 2009; Berne & Stiefel, 1984) have evolved over time from the equity of input measures, such as school resources, to the adequacy of educational outcomes for all students. Concurrently, national education policy has focused on educating *all* students to high standards (e.g., Every Student Succeeds Act (ESSA)). In accordance with these conceptions of equity, we denote schools that have economically disadvantaged students who are achieving at or above the state average in both reading and mathematics. This denotation is one indication of equity within a school and reveals whether economically disadvantaged students in a school have average performance levels that are at least as high as state averages on standardized reading and mathematics assessments. This subgroup analysis provides another way to distinguish between schools, one based on the size of performance gaps between **economically disadvantaged students in a school and all students in the state.**

[1] An exception to this is *U.S. News High School Ranking* produced by the American Institutes for Research (AIR). For details, please see Duhon et al. (2013).

The 2016 methodology consists of a multi-step analysis for both the absolute and relative rankings, and, at each step, the relative ranking accounts for student socioeconomic status while the absolute list does not. The first step was to assess schools' performance within their respective states. For this analysis, we constructed an academic achievement index (AI) based on school average proficiency rates on state standardized tests. We conducted this "threshold" analysis to identify schools that performed above a defined cut point. For the absolute list, we selected schools that were in the 70th percentile or higher in each state's performance distribution (i.e., top 30 percent). For the relative list, we selected schools that performed .5 standard deviation (SD) or higher than the average school in the state with similar proportions of students eligible for free or reduced-price lunch.

Schools that performed above the defined threshold proceeded to the next step—the "ranking" analysis based on the college readiness data. For this step, Westat surveyed schools on both lists to collect data about college readiness indicators (e.g., ratio of students to counselor full-time equivalent (FTE), percentage of students who took advanced placement (AP) tests, and average SAT scores). Using these indicators, we created a college readiness index to rank the schools. For the absolute ranking, we ranked schools by their college readiness index score. For the relative ranking, we ranked schools by their college readiness index score, controlling for socioeconomic status.

The next section discusses relevant literature, including some common critiques, pertaining to school ranking methodologies.

Review of Relevant Literature and Previous Methods

The purpose of the *Newsweek* high school rankings is to identify the top 500 public high schools in the country. Given the diversity among public high schools across the nation, developing one system to rank schools that adequately accounts for heterogeneity among schools presents a challenge. Gladwell (2011) describes this dilemma in his article in *The New Yorker* in which he discusses the opposing forces of comprehensiveness vs. heterogeneity. According to Gladwell, a ranking system can be comprehensive so long as it is not trying to rank units (schools, in this case) that are heterogeneous. Similarly, a ranking system can account for heterogeneity so long as it does not also strive for comprehensiveness. Ranking the nation's top public high schools is, by nature, comprehensive. Yet, there is great diversity among public high schools, much of which cannot be accounted for. Schools vary with regard to their settings—urban, suburban, and rural—and the variety of curriculum and programs they offer. Even with a narrow focus on college preparation, the extent to which schools focus on AP, IB, or dual enrollment varies considerably, and the extent to which high school students take the SAT or ACT varies considerably by state. Schools also enroll students with significantly different background and demographic characteristics, and studies have repeatedly demonstrated that these differences influence achievement in a variety of ways (e.g., Sirin, 2005; Stinebrickner & Stinebrickner, 2003; Sutton & Soderstrom, 1999). To account for at least some of this heterogeneity, we developed two ranking systems.

The first ranking scheme identified and ranked schools by their performance on the academic achievement and college readiness measures. In essence, this ranking scheme identifies the schools with the highest absolute performance on these measures, regardless of influencing factors. A problem with this approach is that it ignores the influence on achievement attributable to student background characteristics. So while this ranking scheme identifies schools with the highest scores using these metrics, estimates of the association between school factors and performance are confounded with contributions made by students' background and family characteristics. That is, it is not clear to what extent student performance is associated with school versus non-school factors. In an effort to address this issue, we devised the relative ranking method to account for student poverty levels.[2] Using the relative method, rankings are based on school performance levels relative to the socioeconomic status of its student body. Since the relationship between poverty levels and student performance is partially accounted for, schools that are ranked highly using this method are not necessarily the same schools that have highest absolute levels of performance.

Another important decision when constructing ranking schemes is which variables to include and how much each variable will contribute to the final rankings (Bobko, Roth, & Buster, 2007; Gladwell, 2011; Webster, 1999). To determine which variables to include, we referred to previous methodologies (e.g., Duhon, Kuriki, Chen, & Noel, 2013; Streib, 2013) and related research (e.g., Grubb, 2009). There is a vast amount of research about the economics of education and education finance that examines the production function of schools (e.g., Grubb, 2009; Harris, 2010) and provides defensible rationales for

[2] This study is not causal and does not make any causal claims about the influence of schools on students. To separate the influence of the school vs. student background characteristics, a causal study would be required. In this case, we are parsing out the association of the school performance with student backgrounds.

factors that should be included or accounted for in a ranking scheme (e.g., financial resources, teacher experience, staff development, college pressure, school problems, family background, and student connectedness). Unfortunately, limited data are available for all schools, and available data exclude many of the recommended factors. Hence, we relied on common measures that have been used in previous rankings (e.g., Duhon et al., 2013; Streib, 2013). These variables include student performance on standardized assessments, high school graduation rates, SAT and ACT scores, performance metrics on AP/IB exams, and college acceptance and enrollment rates. (Specific items are discussed in greater detail below.)

In addition to commonly used items, we included two variables that we derived from research about school performance metrics. The importance of students' engagement with counselors has been shown to positively influence college attendance rates (e.g., Grubb, 2009) and, thus, could be considered one suitable indicator of college preparation. However, due to data limitations, we used a simple school resource indicated by pupil-counselor FTE ratio to reflect this construct. Additionally, we used another variable, referred to as "holding power" (e.g., Balfanz & Legters, 2004; Rumberger & Palardy, 2005), that indicates the dropout and transfer rate of 9th-grade students and can be considered another appropriate indicator of school quality (e.g., Rumberger & Palardy, 2005).

Typically, data for high school rankings are obtained from national or state-level datasets or from self-reported survey data provided by the school or district personnel. Both sources of information have strengths and limitations. Data from national and state sources typically lag by two or three school years (SYs). For this reason, rankings that rely on information from these sources use data that are two years old (e.g., Duhon et al., 2013). For example, 2015 rankings would rely on data collected for SY 2012–13. A primary benefit of using these types of datasets is that there are uniform reporting procedures and data checks in place, resulting in data that are more reliable with regard to consistency and accuracy, even though they are still self-reported by educational personnel. An alternative is to survey schools directly and ask them to self-report the requested data. While this option allows for the more current data to be collected than can be obtained through state or national data files, this method is more susceptible to participation bias and lacks standardized reporting procedures. To address issues stemming from self-reported data, we established procedures to assess the credibility and feasibility of some data. (These procedures are discussed in the data section.)

The unit of analysis for the rankings is at the school level. We used means of student performance indices to indicate school performance. Using average student performance as an indicator of school performance is a common practice, and in fact, most states tie school funding to these metrics (Cobb, 2002; Hall, 2001). This practice of using average student performance as a measure of school performance in rankings like these has been criticized by some for not distinguishing between *student*- vs. *school*-level performance (e.g., Di Carlo, 2013). However, student means are commonly used as an indicator of school performance, for example, in state funding and accountability systems, and to properly address this issue an analysis would need to be based on student-level data and use a multilevel (e.g., hierarchical linear modeling) approach . However, this kind of student-level data is not currently available for all public high schools, nor is it likely to be available in the future for these purposes due to data protection and use restrictions.

Another important element of ranking schemes is determining how to combine the variables. This process entails determining how items could be used for comparative purposes and how much to weight these items to generate a composite score. Methods can involve a single-stage (e.g., Streib, 2013) or multi-stage process (e.g., Duhon et al., 2013). A benefit of using a multi-stage process is that it establishes a minimum performance standard and conducts additional analysis on just the subsample of schools that meet the initial criteria. Additionally, it also allows for the inclusion of student achievement data that need to be assessed within a state. Due to the lack of comparability across state assessments with regard to their difficulty and content focus, we used relative performance within states to determine the minimum threshold, as opposed to including it as part of the ranking criteria. This approach accounts for the nested structure of standardized assessments within states and identifies schools that met the baseline criteria. (See the methods section for additional details.)

There are also multiple approaches to determining a weighting scheme for the variables. Ways of combining data elements into composite scores have been debated by statisticians and researchers for decades (Bobko et al., 2007). Three common procedures for combining items into a composite score are using expert judgmental weighting, equal (or unit) weighting, and regression weighting. Judgmental weighting relies on using theory and professional judgment to establish a weighting scheme. Equal weighting applies an equal amount of importance to each item in the composite. Regression weighting uses a formula based on the relationship between the measures and a criterion to establish weights, although it is only possible to use this approach when there is a criterion measure, limiting its viability in this current case. For public high school rankings, previous methods have relied on judgmental weights (e.g., Duhon et al., 2013; Streib, 2013); however, equal weighting is a viable option. Research that examines judgmental versus equal weighting (e.g., Bobko et al., 2007) points out that, in many cases, each approach produces similar results. Yet, in some instances, it makes more sense to weight items based on theorized substantive (or empirical) importance of the variables. Based on this information, we used a judgmental weighting scheme to generate the composite scores and, subsequently, rank schools

One problem to acknowledge and potentially control for when using judgmental weights is the inter-correlations between the items. In some cases, items can be so similar that they are measuring the same thing. Previous ranking studies (e.g., Duhon et al., 2013; Streib, 2013) do not indicate whether or how multicollinearity among the variables was accounted for. This lack of clarity is problematic because if the items are moderately or highly correlated, the explicit weights may vary substantially from the actual contributions of the items to the rankings (e.g., Webster, 1999). Furthermore, items with very high correlations (e.g., r=+.9) should not all be included in a model. If two items are correlated at r=.9, they are essentially measures of the same underlying construct. Given this, we examined variables for high inter-correlations. If variables were highly correlated, we excluded one to produce a more efficient and simpler model.

The following section explains and elaborates on the 2016 methodology.

Methodology for Identification of Top Schools for *Newsweek*'s Top Public High Schools Rankings

For the 2016 methodology, we conducted a multi-step process consisting of a threshold, a ranking, and an equity analysis. The threshold analysis assessed schools' performance as measured by students' achievement levels on standardized assessments and identified (1) the schools that have the highest levels of academic achievement (for the absolute ranking) and (2) the schools that have the highest levels of academic achievement given the socioeconomic status of their students (for the relative ranking). Schools that were identified in the threshold analysis were surveyed to obtain data about college readiness indicators and, pending completion of the survey, proceeded to the second analysis. The second analysis ranked schools by their responses to several college readiness indicators that identified (1) the schools with the highest levels of college readiness, and (2) the schools that have the highest levels of college readiness after accounting for the socioeconomic status of their students.

The methods used to develop the rankings were designed to accomplish the following:

- Identify high schools within each state that have the highest performance as measured by academic achievement on state assessments in reading and mathematics;
- Assess the extent to which the top-performing schools have prepared their students for college and rank them accordingly.

The procedures are similar for both the relative and absolute rankings; however, for the relative ranking, we accounted for the socioeconomic status of the schools' student body. Details related to each step in the process follow.

Threshold Analysis. We created a high school achievement index based on performance indicators (i.e., proficiency rates on state standardized assessments). For the absolute list, the index was used to identify high schools that perform at or above the 70th percentile within each state. For the relative list, the index was used to identify high schools that perform .5 SD or more than their state's average when accounting for students' socioeconomic status.

Ranking Analysis. For the high schools on both lists identified in the threshold analysis, we created a college readiness index based on six indicators: the ratio of students to counselor FTE, changes in 9th- and 12th-grade student enrollment rates (referred to as holding power), high school graduation rates, a weighted SAT/ACT composite score, a weighted AP/IB/dual enrollment composite score, and the percentage of students enrolling in college. The weighting scheme for the index is shown below:

- Holding power—10 percent
- Ratio of counselor FTE to student enrollment—10 percent
- Weighted SAT/ACT—17.5 percent
- Weighted AP/IB/dual enrollment composite—17.5 percent

- Graduation rate—20 percent
- College enrollment rate—25 percent

For the absolute rankings, we rank ordered the schools by their college readiness index scores. For the relative list, we ranked the schools based on how well the schools performed relative to schools with similar proportions of students eligible for free or reduced-price lunch.

Exhibit 1 provides an overview of the steps, analysis, outcomes, and data used for the 2016 methodology.

Exhibit 1.—Summary of methodology for 2016

Threshold Analysis: Student Achievement	**Ranking Analysis: College Readiness**
Identification of high- achieving schools For relative list: • Accounted for socioeconomic status of students	Identification of schools achieving high marks on college readiness For relative list: • Accounted for socioeconomic status of students
Analysis Creation of academic achievement index (AI) scores for each school by state For relative list: • Within each state, scatterplot of schools' AI scores by socioeconomic status indicator (percent free or reduced-price lunch)	**Analysis** *Of the schools that proceed from the threshold analysis:* Creation of college readiness index scores for each school For relative list: • Scatterplot of schools' college readiness scores by **socioeconomic** status indicator (percent free or reduced-price lunch)
Outcome For absolute list: • Schools within each state were rank ordered by their AI score. Schools in the top 30 percent within their state proceed to step 2. For relative list: • Schools with an AI score of 0.5 SD above the line of best fit selected to proceed to step 2.	**Outcome** For absolute list: • Schools were rank ordered by their scores on the college readiness index. For relative list: • Schools ranked based on their (standardized) distance from the line of best fit.
Source NCES ED*Facts* data Indicators: • Academic achievement in mathematics and reading/language arts	**Source** Email/web-based school surveys, NCES data Indicators: • Ratio of student enrollment to counselor FTE, change in student cohort enrollment rates from 9th to 12th grade (i.e., holding power), percent of students taking SAT/ACT, average SAT/ACT score, graduation rates, college enrollment rates, percent of students taking at least one AP/IB, ratio of AP/IB tests per student, percentage of AP test taken with scores of 3 or higher, average IB scores, and log of dual enrollment credits

Data Sources

The threshold and equity analyses are based on state standardized assessment data obtained from NCES. Assessment data available on states' websites vary with regard to their completeness, limiting comparability and, in some cases, excluding states from the analysis. Hence, we used data from NCES about performance for all public high schools to determine which schools met the initial threshold criteria using systematically collected data for a more complete set of schools. Also, given extensive changes in statewide assessments and/or field-testing during SY 2013–14, we used NCES assessment data from SY 2012–13.[3] Of the schools contained in the NCES data, we selected all public high schools, excluding schools that were closed during the previous report period and those scheduled to open within the subsequent two years, as defined by the NCES data manual. For the threshold analysis, we used the NCES records and data for public high schools. We used this sample for the threshold analysis to establish survey samples for both the absolute and relative lists. Of the 15,819 public high schools, 6,477 schools met the criteria for either list. The threshold analysis identified 4,760 schools for the absolute list and 4,452 for the relative list. A total of 2,735 schools were on both lists (which is 57.4 percent of the absolute schools and 61.4 percent of the relative schools).

We collected survey data from the schools that met or exceeded the threshold criteria on both lists (n=6,477). A link to the web-based survey was sent to the schools via regular mail and by email to collect basic demographic information, graduation rates, college enrollment rates, number of counselor FTEs, number of students taking the SAT/ACT, average SAT/ACT scores, percentage of students taking at least one AP/IB/AICE course, average AP/IB/AICE scores, and dual course credits. We collected data from SY 2014–15, the most recent available data. For the schools that received a survey, the overall response rate was 24.5 percent (n=1,588); the response rate was 21.7 percent (*n*=968) for schools on the relative list and 28 percent (*n*=1,332) for the schools on the absolute list.

We used the survey data to create the college readiness index score for the schools in both of our samples except for the holding power variable, which was calculated as the ratio of survey-reported SY 2014–15 12th-grade enrollment to SY 2011-12 9th-grade enrollment from the NCES Common Core of Data (CCD). However, given the limitations of the data in the CCD, we were not able to account for students who transferred to a different school during this period.

After collecting the data, we ran checks to identify schools with data that were problematic for the ranking scheme. For schools that did not report the number of students taking a test (e.g., SAT or AP), we treated those cases as missing when creating the composite measure, and for schools that reported more AP tests scoring 3 or above than AP tests taken, we treated those cases as missing as well. In addition, we treated the holding power data as missing for any schools with a standardized score above 3 or below -3. When schools reported numbers that resulted in over 100 percent of total enrollment taking an AP exam or graduating, we capped the percentage at 100 percent. For example, we capped responses for all schools that reported having more students taking the SAT or ACT—or more students taking AP or IB classes—than the total number of students enrolled. We also capped the graduation rate at 100 percent for schools

[3] For a discussion of changes in state assessments in reading/language arts and mathematics in SY 2013–14, see U.S. Department of Education. (2016). *State assessments in reading/language arts and mathematics- School year 2012-13 ED*Facts *data documentation.* Washington, DC: Author. http://www.ed.gov/edfacts.

that reported having more high school graduates than total 12th-grade students and the college enrollment rate at 100 percent for schools that reported having more college enrollees than high school graduates. While in some cases it is possible for schools to graduate more students, or have more students enroll in college, than the number of 12th-grade students, capping these two variables at 100 percent eliminated any advantage a school could gain by graduating students outside of the typical four-year window. Additionally, in response to extreme right skew in the dual enrollment data, we standardized the log of the total dual enrollment credits earned, which ensured that the dual enrollment credit variable would not have an outsized effect on the AP/IB/dual enrollment composite variable.

Analysis Details

This section provides more detailed information about the procedures we used to conduct each analysis and to construct both the absolute and relative rankings.

Threshold Analysis: Based on Academic Achievement Indicators

Absolute ranking: Identify high schools that achieve within the top 30 percent on state English language arts and mathematics assessments within their respective state.

Relative ranking: Identify high schools that have higher than expected achievement based on the state-specific relationship between student achievement and percent eligible for free or reduced-price lunch.

Substep 1: Calculate the academic achievement index

As part of the threshold analysis, we created an AI based on proficiency rates on states' reading/language arts and mathematics assessments. We calculated the index scores by taking the weighted average of the proficiency rates on the two assessments. The equation is:

*Weighted avg.=((# taking reading/ language assessment*pct. proficient)+(# taking math*pct. proficient))/ (# taking reading/ language assessment+ # taking math)*

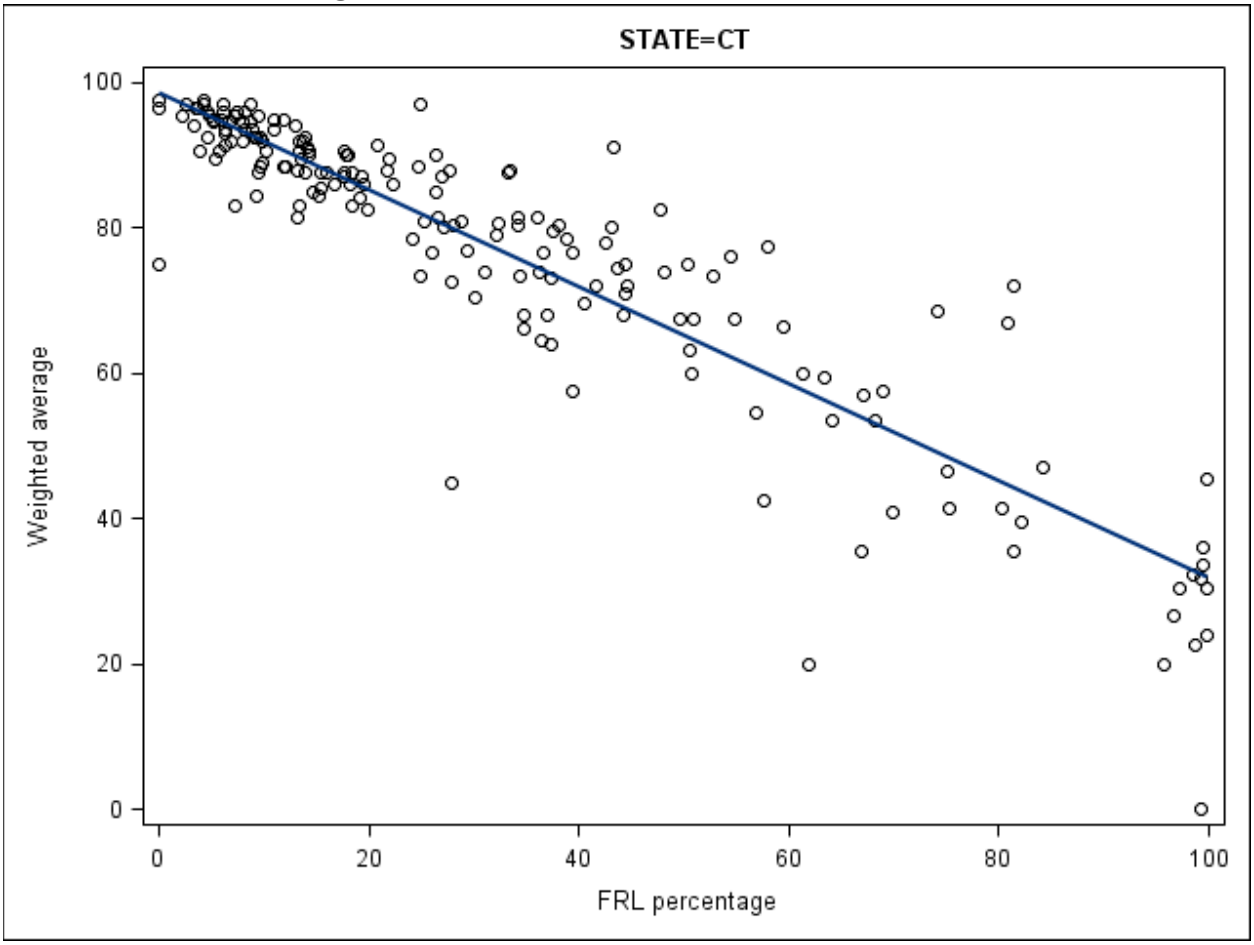
For the relative ranking, we also calculated the percentage of economically disadvantaged students for each high school. To do so, the percentage of students living in poverty for each school was calculated using data from the CCD, which includes the number of students eligible for free or reduced-price lunch and the total number of students in each school. We calculated the percentage eligible for free or reduced-price lunch by dividing the number eligible by the total number of students in the school.

Substep 2: Identify top schools

In this step, for the absolute ranking, we ranked the schools by their AI composite score within state and identified the top 30 percent of schools. Schools in the top 30 percent proceeded to the ranking analysis.

For the relative ranking, we regressed the schools' AI scores on their percentage of economically disadvantaged students to generate a line of best fit that was used to determine the state-specific relationship between the AI and the percentage of economically disadvantaged students. For example, Figure 1 shows the relationship between the index and average student poverty for Connecticut.

Figure 1. Scatterplot of schools' achievement index scores, by percentage of economically disadvantaged students in Connecticut



The extent of the difference between a high school's expected and observed achievement is captured by the residual from the line of best fit. The residual is the vertical difference between the observed value and the expected value indicated by the line of best fit. To identify schools that perform above the average, based on the line of best fit generated by the AI scores and percentage of economically disadvantaged students, we established the cutoff point at +0.5 SD. That is, schools with residuals that met or exceeded +0.5 SD proceeded to the ranking analysis.

Ranking Analysis: Based on College Readiness Indicators

Absolute ranking: Rank high schools by their college readiness index scores.

Relative ranking: Rank high schools by their college readiness index scores, controlling for students' socioeconomic status.

For the ranking analysis, we repeated the procedures used in the threshold analysis, but we used the college readiness indicators to produce a college readiness index score for each school. As in the threshold analysis for the absolute rankings, we ranked schools on their college readiness index score and, for the relative rankings, we ranked schools based on their residuals from the line of best fit using the scatterplots of college readiness index scores by student socioeconomic levels.

Substep 1:Calculation of college readiness index scores for each high school that progressed past the threshold analysis

For the high schools that proceeded beyond the threshold analysis, we created a college readiness index score based on the indicators in the survey. Due to patterns of missing data and high intercorrelations among some indicators, the college readiness index score is based on six indicators: holding power, ratio of counselor FTE to student enrollment, graduation rates, a weighted composite SAT/ACT score, a weighted composite AP/IB/dual enrollment score and college enrollment rates. Regarding high intercorrelations, college acceptance rates and enrollment rates were highly correlated at r=0.97. Due to the high level of this correlation, we only included college enrollment.

To account for instances in which students take both the SAT and ACT assessments, and to account for instances where one or the other test is not typically taken, we created an average weighted SAT/ACT composite. The formula for the weighted SAT/ACT is:

Weighted SAT/ACT=((zSAT# of students)+(zACT * # of students))/(# of students taking the SAT+ # of students taking the ACT)*

Similarly, to account for schools that offer AP, IB, and/or dual enrollment programs, and to adjust accordingly when one program or more programs are offered at a school, we created an average weighted AP/IB/dual enrollment composite. The formula for the weighted AP/IB/dual enrollment composite is:

Weighted AP/IB/Dual Enrollment =((zAP# of students)+(zIB * # of students)+(zLog(Dual Enrollment) * # of students))/(# of students taking the AP+ # of students taking the IB+#of students earning Dual Credits)*

Substep 2: Determination of weighting scheme

We used the same weighting scheme established for the 2014 and 2015 methodology for the six indicators in the 2016 college readiness index. We based the 2014 and 2015 weighting scheme on a review of the literature on previous weighting schemes and the rationales that were provided for using them (e.g., Duhon et al., 2013; Streib, 2013). Finster and Miller (2014) also assessed the empirical relationships among the variables using a principal component analysis. Based on these theoretical and empirical

sources, we created a judgmental weighting scheme that we used to rank the schools. (For a full discussion, please see Finster and Miller, 2014.)

Substep 3: Ranking of schools by their college readiness index scores

Similar to the process used in the threshold analysis, we used the college readiness scores to rank the schools. For the absolute ranking, schools were rank ordered based on their college readiness score. For the relative ranking, we ranked schools by their residuals expressed as the distance from the line of best fit. To do so, we created a scatterplot graph based on each school's college readiness index score and the percentage of economically disadvantaged students. We used these data to generate a line of best fit and to determine a school's college readiness score distance from line of best fit (i.e., the residuals), which we then standardized and used to rank the schools.

Limitations

There are limitations associated with conducting this analysis, many of which stem from the availability of data and their suitability for comparing schools in different states. We have no information about a range of school factors that may influence school performance, such as fiscal resources, teacher quality and effectiveness, school leadership, and school climate. These school factors could potentially contribute to student achievement and college readiness, but data are unavailable for a variety of reasons.

Additionally, the rankings are still dependent on self-reported data, which has implications for our sample and data collection standards. There are circumstances in which the variables may not have been reported consistently. And, this year due to a data collection limitation, the average IB score was rounded to the nearest integer.[4] The "holding power" variable should also be calculated using longitudinal data, however we are limited by using comparisons of cross-sectional data.

Regarding the top high schools, the final rankings reflect only those schools that participated in the survey, not the entire list of schools identified in the threshold analysis. The rankings are dependent on schools that responded to the survey and indicate schools' standings within the sample of respondents. In this sense, both the absolute and relative lists represent the relative ranking of the schools that responded to the survey, not of the entire sample identified in the threshold analysis.

We did not control for any other factor besides student socioeconomic status. Many of the top schools on the relative list are magnet and charter schools and may have an application process that allows them to select high-achieving students. These types of schools would have an advantage in our ranking methodology over schools that do not have a selection process.

In the analysis for the relative list, we used school-level student proficiency rates to select schools that performed better than a predetermined threshold. For example, the relative high school rankings, which controlled for student poverty levels, could be biased against schools with high levels of achievement and

[4] However, while this may affect specific schools, based on a simulation using data from the previous year, the overall impact on the overall rankings was negligible; the rankings derived from IB scores with and without a decimal were correlated at $r>.999$.

college readiness but low levels of student poverty. This bias is, at least partially, the result of ceiling effects on state assessments across schools within a state.

Another limitation is that the school-level proficiency rates in ED*Facts* are a cross-sectional measure and only represent a school's performance at a single point in time. As a result, these data cannot be used to infer causality (i.e., we cannot say for certain that some action or actions taken by a school resulted in a higher ranking than a school that did not take that same action or actions) or make comparisons across states.

For a full discussion of limitations, see Finster and Miller (2014). Future school rankings would still benefit from including more school and non-school factors that have been shown to be associated with student achievement and college readiness to assess school performance and rank schools accordingly.

Conclusion and Discussion

The 2016 methodology predominately follows the 2014 and 2015 methodology by using two approaches to rank high schools. Using the judgmental weighting scheme, the absolute list provides a ranking of schools with the highest average performance on the college readiness index without regard for other factors. The relative list provides a ranking of schools that perform the highest, using the same indicators and weights but after controlling for average student poverty levels. This process controls for non-school factors and attempts to identify the student performance attributable to the school. Control for non-school factors in this manner significantly influences which schools are identified as "top" schools. Each approach provides a different perspective of which high schools are achieving the highest levels of performance based on indicators of academic achievement and college readiness.

References

Baker, B.D., & Green, P.C. (2009) Conceptions, measurement, and application of educational adequacy and equal educational opportunity. In G. Skyes, B. Schneider, & D. Plank (Eds.), *Handbook of education policy research*. New York, NY: Routledge.

Balfanz, R., & Legters, N. (2004). How many central city high schools have a severe dropout problem, where are they located and who attends them? Initial estimates using the Common Core of Data. In G. Orfield (Ed.), *Dropouts in America: Confronting the crisis*. Cambridge, MA: Harvard Educational Publishing.

Berne, R., & Stiefel, L. (1984). *The measurement of equity in school finance*. Baltimore, MD: Johns Hopkins University Press.

Bobko, P., Roth, P., & Buster, M. (2007). The usefulness of unit weights in creating composite scores: A literature review, application to content validity, and meta-analysis. *Organizational Research Methods*, *10*, 689-709.

Cobb, C. (2002*). Performance-based accountability systems for public education*. Unpublished manuscript, The New Hampshire Center for Public Policy Studies.

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J.M., Mood, A.M., Weinfield, F.D., et al. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.

Comrey, A.L., & Lee, H.B. (1992). *A first course in factor analysis* (2nd edition). Hillsdale, NJ: Erlbaum.

Di Carlo, M. (2013, May). *A quick look at "best high school" rankings*. Shanker Blog. Retrieved from http://shankerblog.org/?p=8328

Duhon, C., Kuriki, A., Chen, J., & Noel, A. (2013). Identifying top-performing high schools for the "best high school" rankings. Washington, DC: American Institutes for Research.

Finster, M., & Miller, J. (2014, September). Who's at the top? It depends. . . Identification of highest performing schools for *Newsweek's* 2014 high school rankings. *Newsweek.* Rockville, MD: Westat. Retrieved from http://www.scribd.com/doc/238785187/Newsweek-s-Top-High-Schools-Methodology

Finster, M., & Miller, J. (2015, September). Identification of *Newsweek*'s 2015 top 500 public high school rankings. Rockville, MD: Westat.

Field, A. (2005). *Discovering statistics using SPSS* (2nd edition). Thousand Oaks, CA: Sage Publications.

Gladwell, M. (2011, February). The order of things: What college rankings really tell us. *The New Yorker*. Retrieved from http://www.newyorker.com/magazine/2011/02/14/the-order-of-things

Goldhaber, D. (2002). The mystery of good teaching. *Education Next*, Spring, 1-7.

Grubb, W.N. (2009). *The money myth: School resources, outcomes and equity*. New York: Russell Sage Foundation.

Hall, D. (2001). *How states pursue public school accountability and measure performance of public schools, school year 1998-1999*. Unpublished manuscript, The New Hampshire Center for Public Policy Studies.

Harris, D.N. (2010). Education production functions: Concepts. In D. Brewer & P. McEwan (Eds.), *Economics of education*. San Diego, CA: Elsevier Ltd.

Hucheson, G., & Sonfroniou, N. (1999). *The multivariate social scientist*. London: Sage.

Jolliffe, I.T. (1972). Discarding variables in a principal component analysis, I: artificial data. *Applied Statistics, 21*, 160-173.

Jolliffe, I.T. (1986). *Principal component analysis*. New York: Springer.

Kaiser, H.F. (1974). An index of factorial simplicity. *Psychometrika, 39*, 31-36.

Reardon, S. (2014). Education. In State of the Union: The poverty and inequality report. *Pathways: a magazine on poverty, inequality, and social policy*. Stanford, CA: Stanford Center on Poverty and Inequality. Special issue 2014.

Rumberger, R., & Palardy, G. (2005). Test scores, dropout rates, and transfer rates as alternative indicators of high school performance. *American Education Research Journal, 42*(1), 3-42.

Sirin, S.R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417-453. doi:10.3102/00346543075003417.

Stinebrickner, R., & Stinebrickner, T. (2003). Understanding educational outcomes of students from low-income families. *Journal of Human Resources, 38*(3), 591-617.

Streib, L. (2013, May 6). *America's best high schools 2013: Behind the rankings*. Retrieved from http://www.thedailybeast.com/articles/2013/05/06/america-s-best-high-schools-2013-behind-the-rankings.html

Sutton, A., & Soderstrom, I. (1999). Predicting elementary and secondary school achievement with school-related and demographic factors. *The Journal of Educational Research, 92*, 330-338.

Tabachnick, B., & Fidell, L. (2007). *Using multivariate statistics* (5th edition). Boston: Allyn & Bacon.

Toutkoushian, R.K., & Curtis, T.C. (2005). Effects of socioeconomic factors on public high school outcomes and rankings. *The Journal of Educational Research, 98*(5), 259-271.

Webster, T.J. (1999). A principal component analysis of the *U.S. News & World Report* tier rankings of colleges and universities. *Economics of Education Review, 20*, 235-244.