Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 29 Number 12, August 2024 ISSN 1531-7714

What's in a School Grade? Examining How School Demographics Predict School A-F Letter Grades

Margarita Pivovarova, *Mary Lou Fulton Teachers College, Arizona State University*
Audrey Amrein-Beardsley, *Mary Lou Fulton Teachers College, Arizona State University*
Tray Geiger, *Arizona State University*

A-F school letter grade systems, currently used in 13 states across the United States (U.S.), are one popular version of the systems required by federal policy to help states define, rate, and label school quality every year. In this study, we explored the extent to which such grades assigned to schools, as based on objective measures including students' achievement test scores, may reflect school demographics and other, non-achievement-based school indicators. We found that letter grades do indeed reflect school demographics in a non-random way, thwarting the validity of the inferences to be drawn from states' A-F grade system output, which is critically more important when consequential decisions (e.g., school funding decisions, of pertinence in the state of focus in this study – Arizona) are attached to A-F grade output. More specifically, we found that school demographic composition (e.g., race, free-and-reduced lunch [FRL] eligibility, and English language learner [ELL] status) are strongly associated with school letter grades and the combination of these factors correctly predicts the letter grades received by schools with a 75% accuracy.

Keywords: School Evaluation, School Accountability, Student Achievement, Growth Modeling, Validity, Bias, Educational Policy

Introduction

The first United States (U.S.) federal policy move to hold schools accountable for their students' performances on standardized tests began with No Child Left Behind NCLB (2001), the reauthorization of the Elementary and Secondary Education Act (ESEA) related to federal, test-based accountability policies (U.S., 1965). NCLB required that all states provide evidence their students were achieving adequate yearly progress (AYP) each year, to ultimately reach 100% student proficiency across states by 2014. NCLB also required all states to adopt state-level tests, in the core subject areas of reading/language arts and mathematics, in grades 3-8 and once in high school, and use these tests for accountability policies and purposes.

Almost one decade later, through the former Obama administration's American Recovery and Reinvestment Act (2009) and subsequent *Race to the Top* initiative (2011; see also Duncan, 2009), $4.35 billion in federal stimulus funds were awarded to states in which state leaders were monetarily incentivized to refocus their NCLB-based policy efforts towards teacher- versus student-level, yet still test-based accountability. States receiving federal *Race to the Top* funds were incentivized to use students' test scores for even more consequential purposes, at the teacher-level (i.e., teacher evaluation, termination, and compensation). State leaders also had to adopt such policies if they were to secure waivers excusing them from *not* meeting NCLB's prior 100% proficiency by 2014 goal (Dillon, 2010; Layton, 2012).

The current reauthorization of ESEA - Every Student Succeeds Act (ESSA, 2016) - maintains that states still hold school districts, schools, teachers, and students accountable for student learning and achievement "to effect positive change in our lowest-performing schools, where groups of students are not making progress, and where graduation rates are low over extended" (ESSA, 2016). ESSA also still requires states to measure and report on school performance, akin to how NCLB initially required each state to produce an annual report card that indicated whether schools were succeeding and meeting academic targets (U.S. Department of Education, 2004) via compensatory or conjunctive process each state developed for identifying schools, especially in need of improvement. While all states have complied with this school-level accountability policy mandate, as such, there are currently 13 states actively using A-F grading systems for more consequential purposes (e.g., determining school funding). In this study we investigate the role of student demographics in generating the A-F letter grades ranking using the evidence from one of these states.

A-F School Letter Grades

For the purposes of this study, we defined A-F school letter grades as any state's annual achievement profile required via policy or state statute to help states define and label school quality every year for every public school. The A-F scale is akin to an A, B, C, D, and F grading scale commonly used within classrooms within schools (and higher education), ranked from the highest, A, to the lowest and failing grade, F. Fundamentally, these grades are meant to help parents and members of the public assess and compare schools' performance, and at the same time help states hold schools, and the educators within them, more accountable for increasing student learning, as measured by aggregated student achievement scores over time.

Recent data suggest that 13 states are currently using A-F school letter grade policies and procedures, most of which are Republican or Republican-leaning (Amrein-Beardsley et al., 2022; see also Table 1). The specific A-F letter grade system under analysis in this study comes from the state of Arizona, but it is similar to those in use by the other 12 states, many of which were strongly endorsed by the Foundation for Excellence in Education (FEE). The FEE was launched by former Florida Governor (and brother of former U.S. President George W. Bush who signed into law NCLB) Jeb Bush (ExcelinEd, n.d.a., n.d.b.). ExcelinEd describes itself as a "501(c)(3) nonprofit organization focused on state education reform" and operates on approximately $12 million per year of donations from the Bill & Melinda Gates Foundation, Michael Bloomberg Philanthropies, the Walton Family Foundation, and the Pearson, McGraw-Hill, Northwest Evaluation Association, ACT, College Board, and Educational Testing Service (ETS) testing corporations, among others (ExcelinEd., n.d.b.). The state of Florida was the first state to adopt such a system. However, very little research exists on their efficacy beyond, for example, Amrein-Beardsley and colleagues (2022).

All school accountability systems designed to meet the requirements of ESSA include academic achievement, another academic indicator (almost always student longitudinal growth), English language proficiency, an indicator of school quality and student proficiency, an indicator of school quality and

Table 1. States Using School Letter Grade Accountability Systems and Academic Year of Implementation

State	Implementation Year	State	Implementation Year
Alabama	2013-2014	New Mexico	2012-2013
Arkansas	2012-2013	North Carolina	2013-2014
Arizona	2010-2011	Ohio	2014-2015
Florida	1998-1999	Oklahoma	2011-2012
Indiana	2011-2012 (Old System)	Texas	2015-2016
	2015-2016 (New System)	Utah	2013-2014
Mississippi	2012-2013	West Virginia	2015-2016

student success, and graduation rate (for high schools). Most states combine the various indicators to arrive at an overall determination for each school each year. ESSA requires states to weigh achievement and growth more than the other indicators, but states otherwise have some leeway in how they value the other indicators in their accountability systems.

In theory, the grades and scales used across states' A-F systems seemingly satisfy multiple public interests by, for example, helping parents and educators better and more simply understand school accountability systems, akin to A-F letter grading systems commonly applied to students across U.S. schools. Using A-F letter grade systems, accordingly, help others construct their perceptions about school quality, and distinctively very strong, average, or very weak schools (Jacobsen et al., 2014). However, using school report cards might also shape public perceptions to the extent to which people might view more variation between high- and low-performing schools than might actually be real or authentic (Murray & Howe, 2017). Using such report cards also influences parents' perceptions when evaluating potential schools in which to enroll their children. Given most states have open enrollment policies (Education Commission of the States, 2017), this can also present a challenge given parents may not have information beyond the A-F grades offered, and parents might rely on this measure, sometimes in isolation of other indicators, to make enrollment decisions for their children. Like Schneider et al. (2018) noted, "the information available to 'outsiders' can shape perceptions about organizational functionality, impacting public support for a public good" (p. 5).

The same is true at state public and public policy levels. While student achievement, and growth in student achievement are at the center of all A-F systems, the decisions made by state leaders via these A-F systems also have real implications for the public, how the public perceives and understands schools as public goods, how the public uses their understandings for decision-making purposes, and how public leaders make policy decisions, with consequences often attached to the decisions they make (e.g., school funding) about their states' public schools.

Of related concern is that A-F letter grades, even if comprised of more than increases in students' test scores over time, do not often reveal complete or comprehensive stories about schools and their students (Coe & Brunet, 2006; Jackson et al., 2020; Polikoff et al., 2014). This is especially true when school report cards are more analytical versus holistic in nature, when state analysts often in charge of constructing A-F calculations ignore other factors in their A-F systems (e.g., unique student populations served, programs offered to specialized populations, academic offerings and concentrations, community services offered). Accordingly, critics caution that making policy decisions based on A-F grades is potentially quite dangerous, given that such grades often reflect school demographics (e.g., race, free-and-reduced lunch [FRL] eligibility, and English language learner [ELL] status) than student performance and effectively ignore achievement gaps within and across schools (Adams et al., 2016b; see also Adams et al., 2016a; Murray & Howe, 2017). One consistent finding in the research literature is that schools' ratings or the measures that comprise school ratings are associated with schools' demographics characteristics (e.g., Angrist, 2022; DePaoli, 2014; Toutkoushian & Curtis, 2005; see also White, 1982).

Purpose of the Study

Given this background, as well as the concerns presented in the still-currently limited literature in this area, in this study we aimed to address two research questions: 1) To what degree do student demographics (e.g., race, FRL eligibility, and ELL status) and school factors other than student academic achievement predict existing school letter grades? 2) Which of these non-achievement-based variables are most important in predicting schools' letter grade categorizations? We deemed each of these questions both timely and important, in and of themselves, but also in that findings of this study likely have implications for these 13 states, as well as other states in which state leaders may be contemplating the adoption and implementation of similar state-level educational policies.

Study Context

In this study we analyzed the A-F school grade system from the state of Arizona. Arizona represents an interesting case to study the relationship between school letter grades and student demographics due to its widespread availability of school choice. Previous research demonstrated that school letter grades influence parental enrollment decisions especially in the presence of school choice and that this dynamic leads to an increased school segregation (Chakrabarti & Schwartz, 2013; Hart & Figlio, 2015; Pham et al., 2024). In 2022 Arizona opened an education savings account (ESA) program to all students and became the first state in the nation to expand the universal funded eligibility with flexibility to parents.

The Arizona State Board of Education (SBE) uses an A-F letter grade system to fulfill mandates in state and federal law: a) state law (ARS 15-241) specifies that school performance should be measured using the A-F letter grade system (SBE, 2018), and b) ESSA requires states to measure school performance. The SBE uses the same set of scores to meet the requirements of ESSA and to assign letter grades to schools. For example, an A-school is characterized by a distinguished performance on the statewide assessment, significant student growth, high four-year graduation rates, students on track to proficiency; overall performance is significantly higher than state average; a C-school demonstrates adequate performance but needs improvement on some indicators, such as proficiency, growth or graduation rate, and a F-school experiences systematic failures in proficiency, growth and graduation rates (below 67%) and its performance is in bottom 5% of the state.

Arizona's A-F system measures year-to-year student academic growth, proficiency on English/language arts, mathematics, and science tests, the proficiency and academic growth of English language learners, indicators that an elementary student is ready for success in high school and that high school students are ready to succeed in a career or higher education and high school graduation rates. More specifically, Arizona's K-8 schools are graded based on their scores in four areas that are combined into a single summary measure: a) proficiency on AZMerit (30%), the state assessment; b) student growth on AZMerit (50%); c) English language learners' proficiency and growth (10%); acceleration and high school readiness indicators (10%). Schools can also receive bonus points (up to 5%) for science achievement and special education inclusion. See Table 2 for detailed description of each indicator included in the composite letter grade together with their respective weights in the composite (see also ADE, 202 for more details on each indicator with their respective weights). The score is converted to a letter grade using cut scores set by the SBE which vary from one academic year to another and are based on the overall distribution of the scores. For instance, in 2022-23 academic year the cut scores for grade A were set at 84%, for grade B – at 72%, C – 60%, D – 47%, and F – below 47%.

Growth scores are intended to measure how schools support students' academic growth even if they have not reached proficiency and "reward schools and teachers that accelerate their students' achievement" (SBE, 2018, p. 2). The growth scores are calculated based on quantile regression (so called student growth percentiles, or SGP) and intended to compare the achievement of a school's students to peers that are similar academically; unlike other value-added measures they do not take students' demographic characteristics or subgroup membership into account (Arizona Department of Education (ADE), 2019).

Fundamentally, the grades are meant to help parents and members of the public assess and compare schools' performance, and ultimately help states hold schools, and the educators within them, accountable for meeting higher standards to increase student achievement and improve upon other important indicators of school quality (e.g., graduation rates) over time. School letter grades are made publicly available on ADE public facing website before November 1st of each year. Once letter grades are released, school have two weeks to submit an appeal based on a variety of factors including adverse testing conditions, incorrect data, school or community events. While ADE emphasizes that school letter grade is just one general component that provides information about school and signals about school quality, parents are likely to rely on it when making enrollment decisions.

Methods and Data

Methods

We used discriminant function analysis (DFA) to understand which factors not included in the above computational indicators used to calculate Arizona's A-F grades were predictive of Arizona's schools receiving particular A-F grades. DFA describes the differences between groups (in our case, school letter grades) and exploits these differences in the allocation of classifying observations (schools) of unknown

Table 2. Components of School Letter Grades, Grades K-8 (ADE, 2021)

Indicator	Component	Weight
Proficiency	*ASAA English Language Arts and Math & MSAA English Language Arts and Math* Students' weighted performance on English Language Arts and Math, with Highly Proficient (HP) students receiving the most points.	30%
Growth	*Student Growth Percentiles (SGP) English Language Arts and Math* Students' performance in the prior year on AzM2 and their growth in the current year on AASA compared to their peers.	50%
English Language Learners (ELL)	*Proficiency on AZELLA (English Learner Assessment)* School's percentage of students proficient compared to the K- 8 average ELL proficiency. *Growth on AZELLA (English Learner Assessment)* School's change in performance levels compared to the K-8 average change in performance levels in the current year.	5% 5%
Acceleration/ Readiness	*Grades 5, 6, 7, 8 Math* Increases in highly proficient 8th grade math students and decreases in minimally proficient 8th grade math students *Grade 3 English Language Arts* Decreasing the school's current year minimally proficient percentage compared to prior year or maintaining a low minimally proficient percentage *Chronic Absenteeism* Decreasing the school's current year chronic absenteeism percentage compared to prior year or maintaining a low chronic absenteeism rate *Inclusion of Special Education Students in General Education* Mainstreaming a minimum percentage of special education students into a general education classroom *Improved Growth of Subgroups* Improvement in the school's subgroup scores from the prior year's statewide average or meeting the state's target for the subgroup	10%
Science	For proficiency on AzSCI, schools that tested 95% of their Grade 5 or Grade 8 students may earn 1.5 points for scoring above the statewide average or 3.0 points for scoring well above the statewide average.	Bonus up to 3%
Special Education Enrollment	Schools with a percentage of the statewide average of students enrolled in special education earn 1, 1.5 or 2 bonus points.	Bonus up to 2%

group memberships to each group. In our study, we aim to identify non-achievement related school characteristics that discriminate among school letter grades. In particular, our central focus is on determining what school characteristics or combinations of characteristics create maximum separation among schools received different letter grades, thereby increasing the ability to identify the group which each case most closely resembles. Discriminant function is a weighted combination of factors included in the model which allows to predict the separation between the groups with the highest probability. Depending on the number of groups to separate and associated factors, there might be several discriminant functions which are usually ranked from the highest to the lowest predictive powers. The results of the DFA are presented and interpreted via both discriminant functions and two sets of correlations, canonical and standardized canonical. Canonical correlations are equivalent to the correlation between the output of the discriminant function – a predicted numerical equivalent - and the categories of the dependent variable (Mertler et al., 2021). Standardized canonical correlations are akin to the weights assigned to factors (independent variables). These weights describe the relative contributions of these independent variables to the discriminant function. The higher the standardized correlation in absolute value, the more important is the associated factor in separation between the groups. The conventional threshold value for standardized canonical correlation is .3 (see Pituch & Stevens, 2016). Given the calculated weights and observed numerical values of the school level factors, the discriminant function produces a single score for each school observation, so-called discriminant score. The discriminant score is then used to predict the group membership, letter grades in our case.

Comparison of original group membership with their predicted membership based on included in the model factors suggests how strongly these factors are associated with the original classifications. Consistent with our research question, we selected a group of predictors which were available in publicly accessible data. These predictors are school characteristics and include proportions of students by race and ethnicity, proportions of ELLs, and proportions of students eligible for free-and-reduced lunches (FRL). We also accounted for total school enrollment and class sizes as some of the factors that could be potentially associated with school level of student achievement.

Data

We collected data from two public sources, including the ADE and the Common Core of Data (CCD) made available by the National Center of Educational Statistics (NCES). School letter grades for K8 schools and the component scores are publicly available from the ADE for five academic years between 2016-17 and 2021-22.[1] We merged information on grades and components used to compute these grades (i.e., proficiency points, growth points, and bonus and accelerated points; see ADE, 2021) with information on school demographics from CCD based on school identification numbers. Since data were not consistently reported for some of the variables, we used the most recent five school years (i.e., from 2016-2017 to 2021-2022, removing 2020-2021 since standardized tests were not administered due to the COVID-19 pandemic). Our final analytic sample contained 1,491 school-year observations in a longitudinal dataset of schools where for some of the schools we did not have data for all years in the studied time period. As a result, our data represent an unbalanced across years panel of schools.

Our main variable of interest is a letter grade assigned to a school each year, from A to F. Our data also includes all the components of a letter grade: proficiency points, growth points, bonus and accelerated points (see Table 2). The other variables included in the data set are school demographics: shares of White, Hispanic, Black, Asian, American Indian and Alaska Native, Hawaiian Native and Pacific Islander students; shares of English Language Learners, shares of free and reduced-lunch eligible students. We also have data on school location (rural, suburban, town, and city), total school enrollment, and student-teacher ratio (class size).

Descriptive Statistics

We present a distribution of school letter grades in Arizona by year in Figure 1. The distribution is stable over time with about 30 percent of schools in each of the A, B, or C grade category and about nine percent of schools receiving D grades each year. The two exceptions to this pattern were in 2016-2017, when a higher share of schools received B grades and a smaller share of schools received A grades, and in 2021-2022 (i.e., the first school year that occurred after the

[1] State testing was suspended in 2019-20 because of the COVID-19 pandemic. State assessments were administered in 2020-21 and the results were reported but were not used to issue letter grades. The SBE reported school scores and identified schools with below average levels of performance. Letter grades for 2021-22 were issued in November 2022.

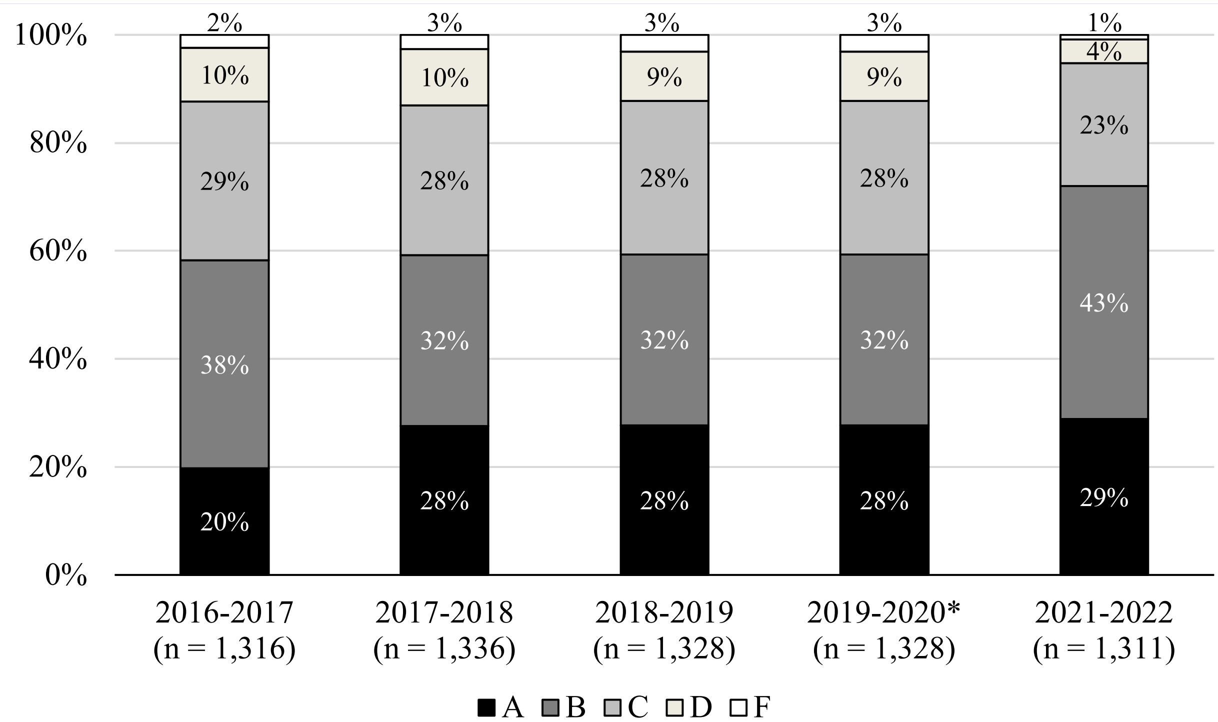
COVID-19 pandemic), when the share of schools receiving D grades dropped to four percent.

In Table 3, which highlights the 2018-2019 school year as an example, we documented average characteristics of schools in each letter grade category and observed some emerging patterns. The most immediate and striking observations, as depicted in Figure 2, were across student demographics. For example, in the 2018-2019 school year, the proportions of FRL-eligible students in C-grade schools (60%) were more than twice that of A-grade schools (28%). A-grade schools had the lowest proportions of Hispanic and Black students, and the highest proportions of White and Asian students compared to schools receiving C through F grades. Schools that received C through F grades had the highest proportions of Hispanic and Black students and the lowest proportions of White and Asian students. Similar patterns held for all other school years.

Higher performing schools (Grade A) were more likely to be situated in urban and suburban areas (81%), and less likely – in rural areas (20%). Failing and low performing schools tended to be smaller compared to A and B grade schools. Schools receiving D grades had on average larger class sizes (25 students vs. 19 students in A schools). The most striking differences between schools with different letter grades are across student demographics.

Following the first stage of the analysis where we found that only first two discriminant functions were significant, we collapsed five categories of letter grades into three categories as following: A and B grades, C grades, D and F grades. We did that to define three group of schools as traditionally perceived by policy and public: high performing schools in categories A and B, average schools represented by C category, and low performing schools in categories D and F. Schools in the last category are more likely to be subject to state

Figure 1. Distribution of Letter Grades, Arizona Public Schools, 2016-2021



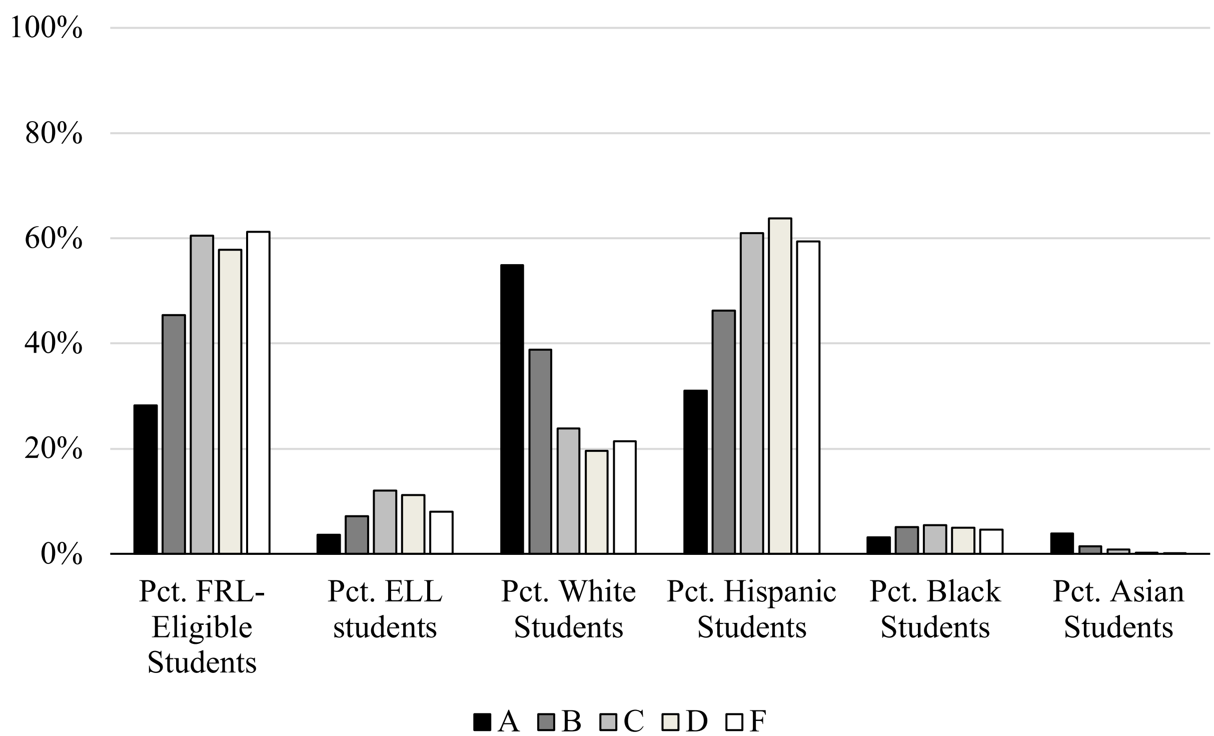
Note 1. Numbers represent shares of school in each letter grade category.

Note 2. A-F letter grades issued for the 2019-2020 school year were the same as those issued for the 2018-2019 school year due to changes in public school operations resulting from the COVID-19 pandemic. For more information, see Arizona Department of Education (n.d.).

Table 3. Descriptive Statistics by Letter Grade, 2018-2019 School Year

	A	B	C	D	F
Urban/Suburban	80.71	75.00	72.95	62.30	48.78
Rural/Small Town	19.29	25.00	27.05	37.70	51.22
Enrollment	586	547	524	493	421
Student-teacher ratio	18.96	18.03	18.94	24.96	20.11
Share of FRL eligible students	28.24	45.41	60.52	57.85	61.16
Share ELL students	3.64	7.19	12.03	11.21	8.01
Share of White students	54.86	38.81	23.85	19.54	21.37
Share of Black students	3.13	5.10	5.43	4.91	4.59
Share of Hispanic students	31.05	46.27	61.01	63.79	59.38
Share of Asian students	3.89	1.39	0.81	0.22	0.10
Growth points	43.51	38.80	33.90	27.82	23.88
Proficiency points	25.47	19.52	14.75	11.58	8.45
N	368	410	377	122	41

Figure 2. Descriptive Statistics by Letter Grade, Student Race, 2018-2019 School Year



interventions should they fail to improve the indicators of performance. With three categories, we estimated two discriminant functions, or two sets of correlations between included variables and group membership (letter grade) represented by one of the three collapsed categories. We present results for each of the discriminant functions in Tables 4 and 5 respectively.

Findings

To answer our first research question – to what extent non-performance related factors could predict school letter grades – we obtained statistics from the DFA. Our model resulted in two discriminant functions, or two sets of different weighted combination of school characteristics which predicted the separation between groups, i.e., letter grades, with a reasonable precision. The first function explained 95% of the variance with a canonical $R^2 = 0.49$. The canonical R^2 in the DFA is similar to an adjusted R^2 in a linear regression. In our case, it means that school level factors that we included in the model as independent variables, were able to capture about 50 percent of the relationship between school grades and these factors taken together. The second function explained the remaining 5% of the variance with a canonical $R^2 = 0.12$. Overall, the two functions accounted for approximately two thirds of the relationship between school letter grades and non-achievement variables which we included in the models. Both functions were statistically significant in separating the schools by grade based on the weighted combination of included factors with the p-value for the first function of less than .001 and second function less than .05. The first function was highly correlated with the outcome variable, $r = .51$.

Of the original grouped cases, about 75% of cases were correctly classified into corresponding grades. The most accurate predictions were observed among the highest performing groups of schools—those that received A and B grades, and the lowest performing group – underperforming and failing schools. We also compared the characteristics of schools which were correctly classified with the ones which were misclassified by our predictive analysis (see Table 6).

Table 4. Discriminant Function One: Canonical Loadings and Standardized Canonical Coefficients

Variable	Canonical Correlation (loadings)	Standardized Canonical Coefficient (weights)
	(A)	(B)
Enrollment	.10	-.03
Student-teacher ratio	-.07	-.14
Share of FRL eligible students	-.84*	-.39
Share ELL students	-.66*	-.11
Share of White students	.84*	1.50
Share of Black students	-.19	.21
Share of Hispanic students	-.77*	1.04
Share of Asian students	.43*	.32
N	1,491	

Note: * indicates the largest absolute correlation between each variable and a discriminant function.

Table 5. Discriminant Function Two: Canonical Loadings and Standardized Canonical Coefficients

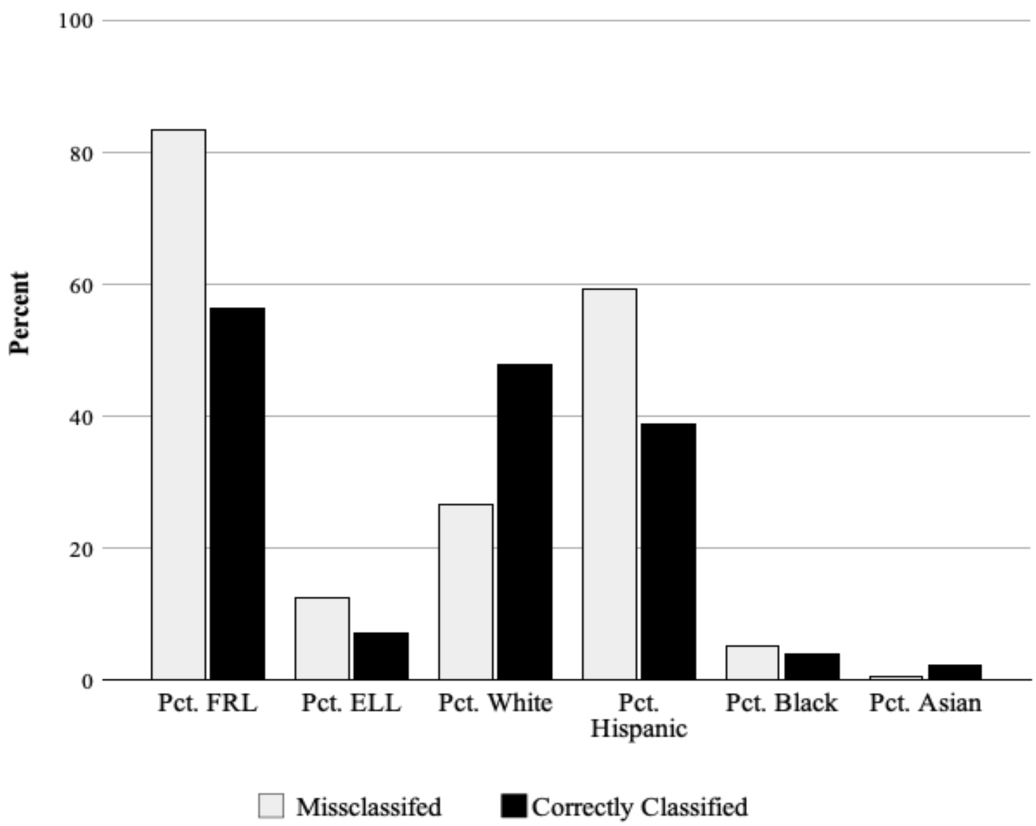
Variable	Canonical Correlation (loadings)	Standardized Canonical Coefficient (weights)
	(A)	(B)
Enrollment	.45*	.43
Student-teacher ratio	.25	.24
Share of FRL eligible students	.31	.78
Share ELL students	.29	.65
Share of White students	.15	1.82
Share of Black students	.09	.14
Share of Hispanic students	-.09	.87
Share of Asian students	.06	.14
N	1,491	

Note: * indicates the largest absolute correlation between each variable and a discriminant function.

Table 6. Comparison Between Correctly Classified and Misclassified Schools by Letter Grades

	Correctly classified schools	Misclassified schools	*P*-value (t-test)
Enrollment	643.69	589.70	< 0.001
Student-teacher ratio	19.82	18.57	0.403
Share of FRL eligible students	0.56	0.83	< 0.001
Share of ELL students	0.07	0.13	< 0.001
Share of White students	0.48	0.26	< 0.001
Share of Black students	0.04	0.05	< 0.001
Share of Hispanic students	0.39	0.59	< 0.001
Share of Asian students	0.02	0.01	< 0.001
Growth points earned	37.16	35.23	< 0.001
Proficiency points earned	21.15	16.76	< 0.001
N	1109	370	

Figure 3. Characteristics of Correctly Classified and Missclassfied Schools



Correctly classified and misclassified schools are statistically different along all characteristics that we included in the model except for student-teacher ratio. The pattern that emerged supports existing research evidence of the strong association between school demographics and achievement. Thus, schools that were more likely to be misclassified, are schools with the higher share of FRL-eligible students, ELL students, Hispanic and Black students, and lower shares of White and Asian students. Visual evidence of that is presented in Figure 3. Stated differently, school grades are likely to reflect school demographics through indicators of growth and achievement.

To address our second research question – which of the non-performance-based indicators are the best predictors of school letter grades – we evaluated the canonical coefficients and canonical loadings. These helped us better contextualize these discriminant functions by identifying the independent variables with the strongest relationships to the discriminant function. As mentioned in the methods section, the canonical correlation (Column A in Tables 4 and 5) is equivalent to the correlation between the output of the discriminant function (i.e., the discriminant score) and the categories of the dependent variable (Mertler et al., 2021). Standardized canonical coefficients (Column B in Tables 4 and 5) describe the relative contributions of independent variables to each discriminant function. The results for the first function suggest that the proportion of White students in a school was the most significant factor in determining a school's grade category, followed by the proportion of Hispanic students. Similarly, for the second function, the proportions of White students held the most weight in predicting schools' A-F group memberships. Overall, school level factors—enrollment and class size—had significantly lower predictive power in group membership, especially as compared to school demographics. The combination of factors together with their respective weights (standardized canonical correlations) can usually be qualitatively summarized and described. In case of our analysis, the functions are capturing the degree of affluence and prevalence of White students in the school's student body. The larger and positive weights are associated with the share of White students and the negative weights with the shares of FRL-eligible students, ELL, and Hispanic students. This means that schools with the high values predicted by the discriminant function are schools with the larger proportion of White students and smaller shares of FRL-eligible, ELL, and Hispanic students. This conjecture is confirmed by the average values of the discriminant function for each of the three letter grade groups. The A and B school on average have a value of the discriminant function of .48, C schools have value of -.65, and D and F school have an average mean of the function equal to -.87. Again, what it means descriptively is that as we move from A to F, we observe higher shares of Hispanic students, ELL students and FRL-eligible students and smaller share of White students.

Discussion and Implications

First, our findings suggest that attributing school failure measured by assigned letter grades to students' lack of effort or other presumed deficiencies is inaccurate and can potentially perpetuate deficit-minded assumptions about school performance. Instead of problematizing students, we should look at how educational, political, and legal institutions may have failed students themselves. It is critical to ask what institutions can do to address the inequities in education access and opportunities which are reflected in the school letter grades. As we demonstrated, school letter grades to a great degree capture student demographics in addition to measures of academic progress.

Second, and related, it is important to underscore that such school ratings are strongly associated and stem from inequities and underinvestment across U.S. public schools often driven by U.S. district and school funding and finance policies and policy-based decisions. As an example, in our data schools receiving D grades had on average larger class sizes (25 students vs. 19 students in grade A schools). This is an unequivocal indicator of resource inequities.

One of the unintended consequences of school letter grades as an accountability tool is the extent to which real estate agents continue to use schools' test scores to gauge neighborhood quality, for example, when purchasing a home. Given the signaling values of letter grades, or student achievement in general, of school quality (Black, 1999; Holme, 2002; Hussain, 2023; Schellenberg & Walters, 2020) this has a potential to lead to segregation of neighborhoods via housing prices, and as a result to more inequities in school funding.

In this study we documented the strong relationship between the A-F grades that schools received in Arizona, as Arizona's primary indicators of student and school quality, and student and school demographics. This finding adds to the numerous existing evidence since Coleman Report (1966) that student background variables (i.e., demographics including race, ethnicity, socio-economic statuses, English language proficiency, special education needs) and other out-of-school factors are significantly more correlated with educational outcomes than in-school resources such as teacher and school quality both at a student and school levels (Cunningham & Sanzo, 2002; Klein et al., 2000; Perry & McConney, 2010; Reardon, 2011; Sirin, 2005; White, 1982). Given that, we acknowledge that the extent to which our findings add to this literature set may not be all that new and exciting. That said, we believe that our study contributes to this literature as we documented the predictive nature of school demographics for school letter grades. Put differently, and in many ways in retrospect, the findings we advance in this study match what we may have predicted from the decades of research on the relationship between demographic factors and student achievement; research that should be considered when developing accountability systems.

Finally, our findings have highlighted the need to develop better, more holistic, and less test-dependent A-F grading systems. Perhaps one option may be for such states to not rely solely on measuring students' test scores one year at a time and, rather, rely more heavily on measuring aggregated levels of school-level growth over time. Another option would be for policymakers to make such systems relying on other indicators that may be more immune to school level demographics and that may more effectively capture what states value in terms of their public education systems (e.g., innovations offerings). Take for example, North Carolina changing its A-F school letter grade system from one to four separate letter grades in: academics, progress, readiness, and opportunity to provide indicators for where and what schools need the most.

Conclusion

For this study, we examined factors beyond those commonly included in the computations of school letter grades. We identified that school demographic compositions (i.e., students' race, FRL, and ELL statuses) can be used to significantly differentiate between and among the A-F grades assigned to schools, and predict the same grade that school received based on achievement indicators, with 75% accuracy. This suggests that school letter grades do not perfectly perform on one of their main tasks to capture school quality for accountability purposes but rather school socio-economic composition. This approach could be detrimental to school improvement as well as to well-being of students and families enrolled in schools that receive failing grades.

Overall, we regard our findings as a cautionary note for states that are currently using or consider using A-F systems. Beyond calling for substantially more research on more states' A-F letter grade systems and policies, and research on issues related to the intersections between education research and policy as we have put forward in this study, we call for at least some reconsideration of such policies until the relationships we observed, also perhaps for other states, are better assessed, mediated, or controlled.

References

Adams, C. M., Forsyth, P. B., Ware, J., & Mwavita, M. (2016a). The informational significance of A-F school accountability grades. *Teachers College Record, 118*(7).

Adams, C. M., Forsyth, P. B., Ware, J., Mwavita, M., Barnes, L. L., & Khojasteh, J. (2016b). An empirical test of Oklahoma's A-F school grades. *Education Policy Analysis Archives, 24*(4). https://doi.org/10.14507/epaa.v24.2127

American Recovery and Reinvestment Act of 2009, Public Law 5, U.S. Statutes at Large 123 (2009): 115-521.

Amrein-Beardsley, A., Geiger, T. J., & Winn, K. (2022). States' performance on NAEP mathematics and reading exams after the implementation of school letter grade accountability policies. Cogent Education, 9(1). https://doi.org/10.1080/2331186X.2022.2063223

Angrist, J., Hull, P., Pathak, P.A., & Walters, C.R. (2022). *Race and the mismeasure of school quality* (No. w29608). National Bureau of Economic Research (NBER). https://www.nber.org/papers/w29608

Arizona Department of Education (ADE). (n.d.). *A-F school letter grades*. https://azsbe.az.gov/parents/a-f-school-letter-grades

Arizona Department of Education (ADE). (2019). *2018 A-F letter grade accountability system: Traditional schools business rules and statistical results.* https://azsbe.az.gov/sites/default/files/media/2017-2018%20A-F%20Business%20Rules.pdf.

Arizona Department of Education ADE). (2021). *Arizona 2021-2022 K-8 A-F school accountability plan.* https://azsbe.az.gov/sites/default/files/2023-08/21-22%2520K-8%2520A-F%2520Plan.pdf

Arizona State Board of Education (SBE). (2018). *FAQ: How Arizona's A-F letter grades for schools work.*

https://azsbe.az.gov/sites/default/files/media/FAQ_9-10_0.pdf.

Black, S. E. (1999). Do better schools matter? Parental valuation of elementary education. *The quarterly journal of economics*, *114*(2), 577-599.

Chakrabarti, R., & Schwartz, N. (2013). Unintended consequences of school accountability policies: Evidence from Florida and implications for New York. *Economic Policy Review*, *19*(1). https://doi.org/10.2139/ssrn.2324231

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

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity.* US Department of Health, Education & Welfare. https://doi.org/10.3886/ICPSR06389.v3

Cunningham, W. G., & Sanzo, T. D. (2002). Is high-stakes testing harming lower socioeconomic status schools? *NASSP Bulletin, 86*(631), 62–75. https://doi.org/10.1177/019263650208663106

DePaoli, J. (2014). *Misleading measurements: How Ohio school ratings foster false comparisons. K-12 Education.* Policy Matters Ohio. https://www.policymattersohio.org/press-room/2014/01/22/ohio-public-school-ratings-foster-false-comparisons

Dillon, S. (2010). Obama to seek sweeping change in 'No Child' law. *The New York Times.*

http://www.nytimes.com/2010/02/01/education/01child.html?pagewanted=all

Duncan, A. (2009). The Race to the Top begins: Remarks by Secretary Arne Duncan.

https://www2.ed.gov/news/pressreleases/2009/07/07242009.html

Education Commission of the States. (2017). *50-state comparison: Open enrollment policies.* https://www.ecs.org/open-enrollment-policies/

ExcelinEd. (n.d.a.). *Foundation for Excellence in Education: About.* https://www.excelined.org/about/

ExcelinEd. (n.d.b.). *Foundation for Excellence in Education: Our donors.* https://www.excelined.org/about/meet-our-donors/

Every Student Succeeds Act (ESSA) of 2016, Pub. L. No. 114-95, § 129 Stat. 1802. (2016). https://www.ed.gov/essa

Hart, C. M., & Figlio, D. N. (2015). School accountability and school choice: Effects on student selection across schools. *National Tax Journal*, *68*(3S), 875-899. https://doi.org/10.17310/ntj.2015.3s.07

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

Hussain, I. (2023). Housing market and school choice response to school quality information shocks. *Journal of Urban Economics*, *138*, 103606.

Jackson, C. K., Porter, S. C., Easton, J. Q., Blanchard, A., & Kiguel, S. (2020). School effects on socioemotional development, school-based arrests, and educational attainment. *American Economic Review: Insights, 2*(4), 491-508. https://doi.org/10.1257/aeri.20200029

Jacobsen, R., Snyder, J. W., & Saultz, A. (2014). Informing or shaping public 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

Klein, S. P., Hamilton, L. S., McCaffrey, D. F., & Stecher, B. M. (2000). What do test scores in Texas tell us? *Education Policy Analysis Archives, 8*(49), 1–22. https://doi.org/10.14507/epaa.v13n36.2005

Layton, L. (2012). Rethinking the classroom: Obama's overhaul of public education. *The Washington Post.* http://www.washingtonpost.com/local/education/rethinking-the-classroom-obamas-overhaul-of-public-education/2012/09/20/a5459346-e171-11e1-ae7f-d2a13e249eb2_print.html

Mertler, C. A., Reinhart, R. V., LaVenia, K. N. (2021). *Advanced and multivariate statistical methods: Practical application and interpretation* (7th ed.). Routledge.

Murray, K., & Howe, K. R. (2017). Neglecting democracy in education policy: A-F school report card accountability systems. *Education Policy Analysis Archives, 25*(109). https://doi.org/10.14507/epaa.25.3017

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115 Stat. 1425 (2002).

Perry, L. B., & McConney, A. (2010). Does the SES of the school matter? An examination of socioeconomic status and student achievement using PISA 2003. *Teachers College Record, 112*(4), 1137–1162. https://doi.org/10.1177/016146811011200401

Pham, L., Cadilla, V., & Asycue, J. (2024). *The unintended consequences of school accountability on racial segregation: Evidence from North Carolina* [Conference presentation]. Annual Meeting of the American Educational Research Asssociation, Philadelphia, PA.

Pituch, K., & Stevens, J. (2016). *Applied multivariate statistics for the social sciences* (6th ed). Routledge.

Polikoff, M. S., McEachin, A. J., Wrabel, S. L., & Duque, M. (2014). The waive of the future? School accountability in the waiver era. *Educational Researcher*, *43*(1), 45–54. https://doi.org/10.3102/0013189X13517137

Race to the Top Act of 2011, S. 844--112th Congress. (2011). https://obamawhitehouse.archives.gov/issues/education/k-12/race-to-the-top

Reardon. S. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G. Duncan &R.Murnane (Eds.) *Whither opportunity? Rising inequality, schools, and children's life* chances (pp. 91-116). Russell Sage Foundation.

Schellenberg, J., & Walters, C. R. (2020). Do parents value school effectiveness?. *American Economic Review*, *110*(5), 1502-39.

Schneider, J., Jacobsen, R., White, R. S., & Gelbach, H. (2018). The (mis)measure of schools: How data affect stakeholder knowledge and perceptions of quality. *Teachers College Record, 120*(6).

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

Toutkoushian, R.K., & Curtis, T. (2005). Effects of socioeconomic factors on public high school outcomes and rankings. *The Journal of Educational Research, 98*(5), 259-271. https://doi.org/10.3200/JOER.98.5.259-271

U.S. (1965). *Elementary and secondary education act of 1965*: H. R. 2362, 89th Cong., 1st sess., Public law 89-10. Reports, bills, debate and act. Washington, U.S. Govt. Print. Off.

U.S. Department of Education. (2004). *No Child Left Behind: A toolkit for teachers.* https://files.eric.ed.gov/fulltext/ED483139.pdf

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91*(3), 461–481. https://doi.org/10.1037/0033-2909.91.3.461

Citation:
Pivovarova, M., Amrein-Beardsley, A., & Geiger, T. (2024). What's in a school grade? Examining how school demographics predict school A-F letter grades. *Practical Assessment, Research, & Evaluation*, 29(12). Available online: https://doi.org/10.7275/pare.2039

Corresponding Author:
Margarita Pivovarova
Arizona State University
Email: margarita.pivovarova@asu.edu