EdWorkingPaper No. 21-415

Free and reduced-price meal eligibility does not measure student poverty: Evidence and policy significance

Ishtiaque Fazlul
University of Missouri

Cory Koedel
University of Missouri

Eric Parsons
University of Missouri

Free and reduced-price meal (FRM) eligibility is commonly used in education research and policy applications as an indicator of student poverty. However, using multiple data sources external to the school system, we show that FRM status is a poor proxy for poverty, with eligibility rates far exceeding what would be expected based on stated income thresholds for program participation. This is true even without accounting for community eligibility for free meals, although community eligibility has exacerbated the problem in recent years. Over the course of showing the limitations of using FRM data to measure poverty, we provide promising validity evidence for a new, publicly-available measure of school poverty based on local-area family incomes.

VERSION: May 2021

Suggested citation: Fazlul, Ishtiaque, Cory Koedel, and Eric Parsons. (2021). Free and reduced-price meal eligibility does not measure student poverty: Evidence and policy significance. (EdWorkingPaper: 21-415). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/ytaf-nc39

Free and reduced-price meal eligibility does not measure student poverty: Evidence and policy significance

Ishtiaque Fazlul
Cory Koedel
Eric Parsons

May 2021

Free and reduced-price meal (FRM) eligibility is commonly used in education research and policy applications as an indicator of student poverty. However, using multiple data sources external to the school system, we show that FRM status is a poor proxy for poverty, with eligibility rates far exceeding what would be expected based on stated income thresholds for program participation. This is true even without accounting for community eligibility for free meals, although community eligibility has exacerbated the problem in recent years. Over the course of showing the limitations of using FRM data to measure poverty, we provide promising validity evidence for a new, publicly-available measure of school poverty based on local-area family incomes.

Acknowledgement
We thank the Missouri Department of Elementary and Secondary Education for access to data, and Andrew Estep and Cheng Qian for research support. We gratefully acknowledge financial support from the Walton Family Foundation and CALDER, which is funded by a consortium of foundations (for more information about CALDER funders, see www.caldercenter.org/about-calder). All opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the funders, data providers, or institutions to which the author(s) are affiliated. All errors are our own.

1. Introduction

Free and reduced-price meal (FRM) eligibility plays a central role in identifying high-poverty students in U.S. education policy. For example, under the federal Every Student Succeeds Act (ESSA), the accountability systems of all fifty states plus Washington DC track gaps in student achievement by poverty status, and 84 percent (all but six states and Washington DC) use FRM data to identify high-poverty students. FRM data are also used to allocate federal, state, and local funding with the goal of targeting resources toward schools serving low-income children.[1] The scholarly community is similarly reliant on FRM data to identify high-poverty students for a variety of research-based applications (Domina et al., 2018).

Noting that FRM data are commonly used in these roles, it is well-understood that FRM designations are error-prone, blunt indicators of poverty and obscure wide variation in income within FRM status bins (Domina et al., 2018; Harwell and LeBeau, 2010; Michelmore and Dynarski, 2017; Parsons, Koedel, and Tan, 2019). There is also evidence that FRM status is awarded to more students than income-eligibility thresholds would imply. For example, Bass (2010) shows disparate trends in youth poverty rates measured inside and outside of schools from the 1970s through the early 2000s. He also provides evidence from district audits showing that for many FRM-designated students, proof of income does not materialize when requested. Domina et al. (2018) link FRM data to IRS tax records and show that FRM status is awarded to more students than the income data suggest should be eligible.

We complement these previous studies by using two new data sources to evaluate the accuracy of FRM data for measuring student poverty. The first data source is administrative records on student Direct Certification (DC) status. DC data capture student participation in social service programs outside of public schools. Participation in these programs is tied to income-eligibility and more carefully vetted than participation in the National School Lunch Program (NSLP). The direct certification criteria in our context—the state of Missouri—are such

[1] According to data compiled by EdBuild, 33 states use FRM data to allocate increased funding toward students from low-income households (e.g., see here, retrieved on 04.11.2021: http://funded.edbuild.org/reports).

that in expectation, the share of DC students in a school should match the share of students at or below 130 percent of the poverty line, which is the stated threshold for *free*-meal eligibility under the NSLP. The second data source is recently-developed "school neighborhood poverty" (SNP) metrics made available by the National Center for Education Statistics (NCES). These metrics are estimated using data on incomes of nearby households for schools and were first made available by the NCES in 2016.

Both of these data sources are independently promising for measuring poverty. We increase our confidence in their reliability by successfully validating them against each other. To perform the validation, we first modify the NCES poverty estimates—which are reported as continuous local-area income measures relative to the poverty line—to produce estimates of the share of students at or below 130 percent of the poverty line for each school, which is the threshold that should correspond to DC status in Missouri. Next, we show that we can accurately predict the share of students who are directly certified, on average, using our modified NCES-based estimates for schools.

After successfully validating the administrative DC data and the NCES poverty data, we apply them to assess the accuracy of FRM data for measuring poverty. Our findings complement previous work by Bass (2010) and Domina et al. (2018) by showing that FRM data overstate poverty; moreover, we provide plausible estimates of the extent to which they do so. Our data additionally allow us to look at free-meal and reduced-meal student designations separately, which reveals the further insight that most of the oversubscription of students in the NSLP is in the "free meal" category. We show that the oversubscription of students in the NSLP is not the result of the NSLP's community eligibility provision (CEP), although the CEP has made it worse.

We then extend our analysis to estimate oversubscription in the NSLP in 27 other states using our NCES-based poverty estimates, which can be constructed for most schools in the U.S. We combine these estimates with FRM eligibility rates from the Common Core of Data (also from NCES). On average, we estimate that the rate of oversubscription in the NSLP nationally is similar to the rate in Missouri, although we also uncover substantial heterogeneity across states. Some of the heterogeneity is surely real—presumably the product of differences across states in districts' FRM certification processes and/or the willingness of families to participate—but some may also be attributable to errors in the Common Core of Data.

Our findings contribute to a thin literature on a topic of great importance for contemporary education research and policy in the United States. We provide the most comprehensive evidence to date that FRM data do not measure poverty in public schools accurately. There are several possible explanations for the persistent overstatement of poverty rates in FRM data, with a prominent one being that districts are incentivized to identify students as FRM eligible but are not similarly incentivized to do so accurately.

A supplementary contribution of our study is that to the best of our knowledge, we provide the first external validity evidence for the NCES' school neighborhood poverty (SNP) metrics. These estimates are publicly available from the NCES for most schools in the U.S. and estimated using a common, central methodology nationwide. Compared to FRM data, which in addition to being inaccurate on average are also subject to different administration and reporting processes across districts and states (see below), the SNP estimates are appealing measures of poverty.

A clear takeaway from our study is that researchers and policymakers should reconsider their interpretation of FRM data as measuring student poverty, even prior to the introduction of community eligibility. FRM data do measure student disadvantage, broadly defined, but not poverty *per se*.[2] This distinction is important in the modern context of the U.S. public education system for two reasons. First, there are high stakes attached to poverty measurement in U.S. schools for funding and accountability purposes, which implies that measurement ambiguity can have important consequences for districts, schools, and ultimately students. Second, the common acceptance of the idea that FRM data have historically served to measure poverty accurately is

[2] Domina et al. (2018) provide concrete evidence on this point; also see Harwell and LeBeau (2010) for additional discussion of the distinction between measuring poverty and disadvantage.

hampering the use of modern data systems to develop new and more accurate poverty and disadvantage metrics. For example, old rates of FRM eligibility are being used as benchmarks for assessing the accuracy of new poverty metrics (e.g., see Croninger, Rice, and Checovich, 2015; Grich, 2019; Massachusetts Department of Elementary and Secondary Education, 2017). However, our research shows the FRM-based rates are incorrect, which makes these benchmarking procedures problematic.[3] In addition, as data systems evolve in ways that facilitate the development of more comprehensive measures of "student disadvantage," broadly defined, there is resistance to the idea of moving away from a basic poverty designation and toward the more nebulous—but arguably more useful—concept of disadvantage. However, as our results make clear, FRM data have always captured the more nebulous concept of student disadvantage, albeit under the guise of measuring poverty and without the benefit of using all available information to measure disadvantage more effectively.

2. Background

FRM eligibility for individual students under the NSLP is determined by school districts. Districts assess eligibility in two ways. First, students can be "directly certified" for free meals if they participate in a qualified federal assistance program such as the Supplemental Nutrition Assistance Program (SNAP), Temporary Assistance for Needy Families (TANF), or the Food Distribution Program on Indian Reservations (FDPIR). In addition, foster, migrant, runaway, and homeless youth—and in some states, additional income groups like students eligible for Medicaid—can also be directly certified (Gindling et al., 2018; Greenberg, 2018; Greenberg, Blagg, and Rainer, 2019).

Second, school districts administer income surveys to parents and students can be classified as eligible for free or reduced-price meals based on the survey responses. Students

[3] A related policy complication is that in some states that use FRM data to measure poverty, CEP-adopting schools and districts that no longer collect FRM data are being forced to report poverty using alternative metrics. If these schools and districts do not build measures that match FRM-based poverty rates, it can put them at a disadvantage in state funding and other policies, even if the FRM-based rates are not correct. Gindling et al. (2018) provides a useful case study in Baltimore City Public Schools.

from families with incomes at or below 130 percent of the federal poverty line are eligible for free meals, and those from families with incomes between 130 and 185 percent of the poverty line are eligible for reduced-priced meals. Districts are incentivized to encourage and approve parent applications for the NSLP because they receive meal subsidies for FRM-eligible students and can gain access to additional federal, state, and local funding. Parents' incentives are also aligned—they benefit because participation in the NSLP lowers the cost of food for their children.

Only a very small fraction of NSLP applications go through an income verification process (Bass, 2009).[4] In fact, according to the USDA's Eligibility Manual for School Meals for 2017, attempting to verify more than three percent of applications without special cause is prohibited.[5] In instances where income-eligibility cannot be verified—which is quite common, up to 50 percent of applications (Burghardt, Silva, and Hulsey, 2004)—the consequences are modest. Specifically, FRM status is cancelled, but there are no other repercussions. The incentive structure is clearly in favor of districts and parents stretching the boundaries of eligibility. We do not take a normative stance on whether this is "good" or "bad" from a policy perspective; but for the purpose of relying on FRM data to measure poverty accurately, the incentive structure is cause for concern.[6]

Figure 1 updates a similar figure in Bass (2010) using data from the Digest of Education Statistics through 2018 (de Brey et al., 2021). It plots the national share of FRM-eligible students and the share of school-aged children living at or below the poverty line. The former data are

[4] See NSLP Verification Toolkit from USDA. Retrieved on 03.30.2021 from https://www.fns.usda.gov/cn/verification-toolkit

[5] The 2017 version of USDA's Eligibility Manual for School Meals says, "With the exception of verification for cause, LEAs must not verify more or less than the standard sample size or the alternate sample size (when the alternate sample size is used). LEAs must not verify all (100 percent) of the applications." Verification for cause can be performed if "the LEA is aware of additional income or persons in the household." This information was retrieved from the following address on 03.31.2021. https://fns-prod.azureedge.net/sites/default/files/cn/SP36_CACFP15_SFSP11-2017a1.pdf

[6] Research shows that providing all students in a school with free meal leads to better test scores (Ruffini, forthcoming; Schwartz and Rothbart, 2020), improved student discipline (Gordon and Ruffini, forthcoming), and no increase in the BMI or the probability of being obese or overweight (Davis and Musaddiq, 2019; Schwartz and Rothbart, 2020).

collected by school districts as described above; the latter are based on data from the U.S. Census.[7] The income thresholds corresponding to these definitions are different—i.e., the stated FRM-eligibility threshold is at 185 percent of the poverty line—which limits comparative inference to some degree. Still, the differential trends in the two poverty measures over the 2006-2018 period are suggestive of a measurement problem. Most notably, whereas the share of children in poverty as measured by the U.S. Census moves with the business cycle as anticipated and increases by just 1.5 percentage points from 2006-2018, the FRM-eligible share rises throughout the sample period and increases by more than 10 percentage points over the same period.

The most closely-related study to our own is Domina et al. (2018), who merge FRM eligibility data from Oregon and a single school district in CA with family income data from the IRS. These authors find disagreement in the data in both directions—i.e., seemingly FRM-eligible students based on income who are not enrolled in the program and income-ineligible students who are enrolled. Consistent with our findings below, the latter are much more prevalent than the former. A factor that differentiates our work is that Domina et al. (2018) use highly-restricted IRS data to examine the income-eligibility question. These are excellent data but unlikely to be commonly available to researchers or education agencies. A positive attribute of our study is that our key validation metrics—Direct Certification data and the NCES' SNP measures—are more readily accessible. Access to the former is uncommon but not as difficult to obtain as access to IRS data. The SNP data from the NCES are publicly available.

Finally, the Community Eligibility Provision (CEP) was implemented in the NSLP starting with the 2014-15 school year. The CEP allows schools and districts to provide free meals to all students if the student body is sufficiently impoverished, and in many states, FRM-eligibility data are overwritten for CEP schools to indicate that 100-percent of students are free-meal eligible (Chingos, 2016; Greenberg, Blagg, and Rainer, 2019; Koedel and Parsons, 2021).

[7] These data are reported across several issues of the Digest of Education Statistics, the most recent of which is de Bray et al. (2021).

As a result, the CEP further degrades the link between student poverty and FRM eligibility. In the analysis that follows we show that the CEP contributes to the overstatement of poverty in modern FRM data, but it is not the primary driver and even in the absence of the CEP, FRM data still greatly overstate poverty.[8]

3. Data

3.1 Missouri administrative data

Our primary empirical analysis is conducted using administrative student records from the Missouri Department of Elementary and Secondary Education (DESE) for all students enrolled in public schools during the 2015-16 and 2016-17 school years (school years are hereafter identified by the spring year—e.g., 2016 for 2015-16). We restrict our analysis to schools with at least 25 students. The most important variables in the core administrative data are students' "free" and "reduced-price" meal designations (FM and RM, respectively). Through DESE, we also have access to merged administrative data indicating whether each student is directly certified to receive free meals. We will refer to this combined dataset as the Missouri administrative data.

Students from households that participate in SNAP, TANF, and FDPIR, and students classified as migrant, runaway, homeless, or in foster care are categorically eligible for free meals in Missouri.[9] DESE has an agreement with Missouri Department of Health and Senior Services to provide the program-participation information necessary to directly certify these students. All Missouri districts are required to download direct certification information for their students at least three times annually to make sure all students eligible for FM through direct certification are extended the benefit. Missouri's direct certification processes are above-average among states along several measurable dimensions (Koedel and Parsons, 2021).

[8] One might worry about the impact of the CEP on the FRM trend in Figure 1, but no impact is visually apparent. There are two reasons for this: (1) some states have not overwritten their individual FRM-eligibility data, dulling its impact on a national scale, and (2) even in states in which the CEP has overwritten the data, only a small fraction of the total student population changes FRM designations due to the CEP (Koedel and Parsons, 2021).
[9] Information retrieved from the following address on 03.24.2021: https://dese.mo.gov/sites/default/files/FNS-FreeandReduced-DirectCertbooklet2020-21.pdf.

A feature of the direct-certification landscape in Missouri that facilitates our analysis is that these criteria should identify students living at or below 130 percent of the poverty line. This is useful because this is the same income threshold for free meal eligibility under the NSLP. The most important direct-certification criterion in this regard is SNAP participation. SNAP is the largest program that leads to direct certification in Missouri and it uses the 130-percent-of-the-poverty-line cutoff to determine eligibility.

To be precise, we require the following assumptions to hold in order for the share of students who are directly certified in our data to reflect the share of students who are living at or below 130 percent of the poverty line:

1. Income-eligibility requirements for the social-service programs that lead to direct certification include students up to 130 percent of the poverty line and are strictly enforced.
2. All eligible families participate in social-service programs that lead to direct certification.

It seems implausible that these assumptions are never violated (e.g., surely eligible families exist in Missouri that are not participating in a social-services program that leads to direct certification). However, if violations are modest, the DC data will effectively measure the fraction of students living at or below 130 percent of the poverty line. We rely on a validation exercise to assess the potential magnitude of violations, as described below.[10]

Table 1 provides descriptive information about our sample.

3.2 NCES School Neighborhood Poverty data

Beginning in 2016, the NCES began reporting school neighborhood poverty (SNP) metrics for nearly every school in the United States.[11] These metrics are based on household

[10] A notable contextual feature of Missouri is that the share of the population that is Hispanic is well below the national average. Available research suggests that assumption 2 is likely to be violated in states/locales with large Hispanic population shares (Lichter et al., 2015; Sandstrom, Huerta, and Loprest 2014; Williams, 2013; Zedlewski and Martinez-Schiferl, 2010), but empirically, the results that follow indicate that under-participation is not problematic in Missouri.

[11] For example, in Missouri, 2,172 out of the 2,215 public schools have corresponding SNP metrics from NCES in 2016 (98 percent). In 2017, SNP metrics are available for 2,186 out of 2,219 schools (99 percent).

income data from the U.S. Census Bureau's American Community Survey (ACS). SNP metrics are reported as continuous variables and aim to measure the average income-to-poverty ratio (IPR) in a school, multiplied by 100. A value of exactly 100, for instance, indicates that the average income value is at the poverty line. A value of 200 indicates the average income is double the poverty line, and so on. The IPR metrics are described in Geverdt (2019) as capturing "economic conditions of neighborhoods where schools are located." More specifically, the IPR estimates represent the income-to-poverty ratio in a household that would hypothetically be situated in the exact geographic location of the school.

We elaborate briefly on the construction of the SNP metrics here and refer interested readers to Geverdt (2019) for more information. The SNP metrics are estimated using a spatial estimation process called Kriging. This method uses the weighted sum of income values in measured locations to predict values in unmeasured locations (Cressie 1989; Cressie 1993). The predicted value of an unmeasured location is estimated using the following equation (Geverdt and Nixon, 2018):

$$\hat{Z}(s_0) = \sum_{i=1}^{\mathbb{N}} \lambda_i Z\,(s_i) \tag{1}$$

where $\hat{Z}(s_0)$ is the predicted income value in the unmeasured location, $Z(s_i)$ is the value at measured location i, and λ_i is a weighting parameter. The closer that measured location i is to the unmeasured location of interest, the larger is λ_i. The value of λ_i also depends on the covariance structure of all measured locations; i.e., the relationship between distance and income elsewhere in the data, which is modelled using a semi-variogram when calculating SNP (Geverdt and Nixon, 2018). The NCES IPR estimates for each school are based on data from the 25 households closest to the school in the American Community Survey from the U.S. Census.[12]

[12] An issue with basic Kriging is that it assumes the relationship between the variance of the measure and distance between locations is the same throughout the sample. But that may not be true for SNP estimation conducted on a national scale covering a variety of regions and regional contexts (e.g., urban versus rural areas). NCES's SNP metrics are estimated using empirical Bayesian Kriging which addresses this problem by dividing areas into smaller regions and developing models for each region (Geverdt and Nixon, 2018). The local models take into account differences in spatial dependence across regions.

While conceptually appealing, whether the SNP metrics are accurate measures of poverty is unclear. Sources of potential discrepancies include the following:

1. Not all students who live near a local public school will attend that school. This will be especially problematic if selection into a school is based on income—e.g., if wealthier families in an area with high income heterogeneity send their children to private schools, this would lead to a systematic discrepancy between the incomes of children who attend the public school and the IPR value.
2. Estimated poverty at the exact location of a school may not reflect poverty in a school's catchment area. If homes very close or far from schools are systematically more valuable within catchment areas, the estimates could be biased.

Idiosyncratic circumstances such as these surely exist in the IPR data, but whether they are common or systematic is uncertain. Our validation of the IPR and DC data against each other provides a high-level test of the accuracy of the information in both data sources. To the best of our knowledge, the accuracy of the SNP metrics has not been tested previously —in the process of using these metrics to assess the informational content of FRM data, we provide the first evidence on their validity.

The bottom panel of Table 1 provides basic summary statistics using various poverty measures for Missouri schools in 2016 and 2017. This includes the IPR estimate, which is the direct value reported as the SNP metric by NCES, along with two modified versions of the IPR estimate—IPR(130) and IPR(185)—that we describe in the next section.

4. Methods

We begin by validating the DC and SNP data against each other. To facilitate the validation—and again noting that the DC share in each school is a plausible measure of the share of students living at or below 130 percent of the poverty line—the first step is to manipulate the continuously-measured IPR data to recover an estimate of the share of students living at or below this threshold. Our manipulation of NCES-reported IPR values relies on the assumption that they are mean values from a normal distribution. Along with their standard errors (also reported by

NCES), we can use the IPR estimates to construct the distribution of income in each school under the normality assumption. Then, the fraction of the student population with income values at or below any threshold value can be calculated directly from the cumulative distribution function. Equation (2) gives an example at the focal value of 130 percent of the poverty line:

$$\widehat{IPR(130)}_{jt} = P\ (IPR_{jt} \leq 130) = \int_{-\infty}^{130} f(IPR)dIPR \tag{2}$$

In the equation, $\widehat{IPR(130)}_{jt}$ is the estimated fraction of students at school *j* in year *t* with family incomes at or below 130 percent of the poverty line and $f(IPR)$ is the probability density function (pdf) of IPR.[13] The general form of equation (2), where X indicates a generic income value as a percent of the poverty line, can be written as:

$$\widehat{IPR(X)}_{jt} = P\ (IPR_{jt} \leq X) = \int_{-\infty}^{X} f(IPR)dIPR \tag{3}$$

With estimated $IPR(130)_{jt}$ values in hand, we validate them against schools' DC shares using the following univariate regression, weighted by school enrollment:

$$DC_{jt} = \beta_0 + \widehat{IPR(130)}_{jt}\beta_1 + \varepsilon_{jt} \tag{4}$$

In equation (4), DC_{jt} is the share of directly-certified students in school *j* in year *t* and $\widehat{IPR(130)}_{jt}$ is the estimated value from equation (2). If both variables in this regression are measuring the same construct, on average, then the expected value of β_1 is 1.0. Deviations from 1.0 would imply systematic differences in what the two variables measure. Note that the empirical Bayesian Kriging procedure used to construct the original IPR variables embeds shrinkage, so attenuation bias in β_1 is not a concern (Chetty, Friedman, and Rockoff, 2014; Jacob and Lefgren, 2008).

A sufficient condition for recovering a value of $\beta_1 = 1.0$ is that the assumptions outlined in the previous section for each measure to recover the fraction of students at or below 130

---

[13] The IPR standard errors may understate the variance of the students attending that school because it takes into account only the 25 closest households to the school (with school-age children) and there may be more income heterogeneity in the wider catchment area. Theoretically, this smaller standard error would lead us to underestimate the share of students at or below 130 percent of the poverty line using the procedure outlined in the text. However, a sensitivity analysis omitted for brevity shows this concern is not of practical importance.

percent of the poverty line are satisfied; or, at least satisfied to a rough approximation. Further, given that the assumptions for each measure are very different substantively, it would be highly unlikely to recover an estimate of $\beta_1 = 1.0$ if either or both sets of assumptions are violated, in which case a divergence of the measures seems almost assured. Thus, the test from equation (4) of the null hypothesis that $\beta_1 = 1.0$ is a credible test of the plausibility of the assumptions that underlie *both* measures. If the measures agree, it is difficult to construct a story by which they are both wrong but the sources of errors just happen to align such that they are wrong in the same way and to the same degree.

Below we show that we fail to reject the null hypothesis that $\beta_1 = 1.0$ in equation (4) with a fairly precise confidence interval, which implies that both measures are accurate indicators of the share of students living at or below 130 percent of the poverty line, on average. Taking this as a point of departure, we then estimate the following univariate regressions, also weighted by school enrollment:

$$FM_{jt} = \delta_0 + \widehat{IPR(130)}_{jt}\delta_1 + e_{jt} \tag{5}$$

$$FM_{jt} = \gamma_0 + DC_{jt}\gamma_1 + u_{jt} \tag{6}$$

In equations (5) and (6), we regress the share of students eligible for FM in school *j* and year *t*, FM_{jt}, on the school's IPR(130) estimate and DC share, respectively. By rule, students identified as eligible for FM should include only those in households at or below 130 percent of the poverty line. Therefore, the same logic from equation (4) applies—we should anticipate that δ_1 and γ_1 have values of 1.0. Values above 1.0 would indicate that more students are designated for free-meal status than income-eligibility alone would dictate.[14] Our validation exercise suggests that these two equations are redundant—i.e., $\delta_1 \approx \gamma_1$—but we estimate both for completeness.

We also extend equation (5) to look at the free *and reduced-price* meal eligibility threshold, which is at 185 percent of the poverty line, using equation (7):

$$FRM_{jt} = \lambda_0 + \widehat{IPR(185)}_{jt}\lambda_1 + \eta_{jt} \tag{7}$$

[14] There is some nuance to this interpretation—see Domina et al. (2018) for a discussion.

In equation (7), FRM_{jt} is the share of students eligible for free or reduced-price meals, and $\widehat{IPR(185)}$ is the income-aligned measure based on the SNP data. λ_1 takes on the same interpretation as δ_1 and γ_1 above—i.e., values above 1.0 indicate oversubscription of free and reduced-price meal eligibility. A caveat to this extension is that while it is motivated by the validation regression in equation (4), we do not have a comparable measure to directly validate our estimates of $IPR(185)$. We must assume that the validation of $IPR(130)$ also implies that $IPR(185)$ is an accurate measure of the fraction of families living at or below 185 percent of the poverty line. This assumption is reasonable, but we have no way of providing direct evidence to confirm or refute it.

Finally, we return to the point from above that like many other states, the FRM data in Missouri are affected by the CEP such that some high-poverty schools are coded as entirely comprised of FM students even when income-eligibility is below 100 percent. One could interpret the CEP as "biasing" upward the estimates in equations (5) and (7), although in our view the term "bias" is not appropriate because the CEP is a true source of inaccuracy of modern FRM data. Still, we assess the impact of the CEP on the estimates in equations (5) and (7) by imputing the FM_{jt} and FRM_{jt} values for CEP schools to their values during the last year prior to CEP implementation in Missouri, which was 2014. If the CEP were solely responsible for the oversubscription in FRM data, we would expect our models based on the CEP-adjusted data to recover coefficients on the key parameter of interest of 1.0.

5. Results

Table 2 shows the results from our baseline regressions in equations (4), (5), (6), and (7). The column headers indicate the dependent variable in each model and the year for which the model is estimated (either 2016 or 2017, which are the first two years that SNP metrics are available from NCES).

First, columns (1) and (2) report on the validation regressions of the DC share on IPR(130). We cannot reject the null hypothesis that $\beta_1 = 1.0$ and our confidence intervals are precise. Next, columns (3)-(6) show regressions of the FM share on IPR(130) and the DC share, respectively, as shown in equations (5) and (6). If we believe that student FM eligibility is allocated following income rules, we should also get coefficients of 1.0 in these regressions, but our estimates are much larger. The range of coefficient values is from 1.37-1.51, implying an oversubscription rate in FM eligibility of 37-51 percent. In all cases in columns (3)-(6), we can comfortably reject the null hypothesis of a 1.0 coefficient.

We make two additional observations about the estimates in columns (3)-(6). First, the coefficients on the DC-share variables are somewhat smaller than the coefficients on IPR(130). We did not explore this result in great detail, but note that modest differences along the lines of what we find are not ruled out by the results in columns (1) and (2). This is because those results show that DC and IPR(130) provide the same information about poverty on average, but they do not rule out distributional differences in the variables that could contribute to small differences in the coefficients in columns (3)-(6).[15] The second noteworthy observation is that the standard errors in the IPR(130) regressions are much larger, reflecting greater imprecision in these estimates relative to the DC shares based on the Missouri administrative data.

In the last two columns of Table 2, we extend the analysis to look at the 185-percent-of-poverty income threshold using the FRM share and our IPR(185) measure. These results also indicate program oversubscription, with coefficients in 2016 and 2017 of 1.385 and 1.396, respectively. These estimates are most comparable to the IPR(130) estimates in columns (3) and (4) because they use the same measurement mode. Inference based on both sets of estimates suggests that the *free* meal oversubscription rate exceeds the *reduced-price* meal oversubscription rate. Note that the gap in the oversubscription rate between these levels is understated by a direct comparison of the coefficients because the estimates in columns (7) and (8) are inclusive of the free-meal oversubscription effect in the earlier columns. The finding that

[15] We document and discuss the relationship between the DC share and IPR(130) throughout the distribution in Appendix A.

oversubscription in RM data is less than in FM data is alluded to in our descriptive statistics in Table 1, which show that few students are listed as eligible for reduced-price but not free meals.

Next, we consider the possibility that the CEP is driving the oversubscription of FM and FRM eligibility in our data. Specifically, the concern is that since all students at CEP schools are coded as eligible for free meals in Missouri, this will mechanically increase the coefficient estimates in Table 2 even if FRM eligibility has been a historically accurate measure of poverty (i.e., prior to the CEP). To assess the impact of the CEP, we build a modified dataset in which the FM and FRM shares for non-CEP schools are left as reported in 2016 and 2017, but for CEP schools these values are re-coded to the last pre-CEP year in Missouri: 2014. Table 3 shows results from estimating our models on this dataset. For brevity we show results using the IPR variables only (i.e., we do not report the results using the DC share for brevity, although the results are substantively similar following from Table 2).

The coefficients in Table 3 decline in both the FM and FRM regressions compared to their analogs in Table 2. Specifically, for the regressions of schools' FM shares, the coefficients on IPR(130) from 2016 and 2017 are 1.354 and 1.318, respectively, and for the regressions of schools' FRM shares, they are 1.293 and 1.308. These estimates still imply substantial and statistically significant oversubscription in the NSLP, although at a rate that is about 20-30 percent smaller than what is implied by the analogous estimates in Table 2. We conclude from these results that the CEP can explain some of the oversubscription in modern FRM data, but most of the oversubscription is not attributable to the CEP. This finding is consistent with Domina et al.'s (2018) pre-CEP analysis of FRM data and Koedel and Parsons's (2021) investigation of the scope for impact of the CEP.

Finally, we also test the accuracy of FM data using a framework based on the mean squared error (MSE). For this analysis we make the assumption that our administrative DC data from Missouri are accurate measures of schools' poverty shares, then calculate the MSEs of IPR(130) and the FM share relative to the DC share. Consistent with the analysis outlined thus far, Appendix Table B1 shows that the MSE of FM data is more than double the MSE of

IPR(130). One reason this extra analysis is useful is that the MSE accounts for both bias and precision in the poverty measures. The results in Table 2 show that the IPR estimates are noisier than their DC-share counterparts, but the substantive consequences of this are not obvious. The MSE results make clear that if we assume the DC shares are accurate measures of school poverty, then the IPR(130) estimates are much more accurate than their FM-based analogs.

6. Extensions

6.1 Using student poverty metrics to predict student achievement

In this section we use the different student-poverty metrics to predict student achievement in Missouri. Specifically, we estimate the following cross-sectional regressions at the school level using data from 2017 (results from 2016 are very similar and omitted for brevity):

$$Y_j = \phi_0 + P_j\phi_1 + \varepsilon_j \tag{8}$$

In equation (8), Y_j is the average standardized math test score for students in school j and P_j is a measure of the share of students at 130 percent of the poverty line or below. We estimate this regression three times, where P_j represents either the DC share, IPR(130) estimate, or FM share. The sample includes all schools in Missouri with at least one grade in the 4-8 range (and we continue to impose the condition that enrollment is at or above 25). It is well-documented that student poverty—despite its imprecise measurement using FRM data—is a strong predictor of low student achievement (Domina et al., 2018; Koedel and Parsons, 2021; Michelmore and Dynarski, 2017). Thus, we expect ϕ_1 to be negative in each version of equation (8). Moreover, if all three metrics are capturing the same information about poverty, we would expect ϕ_1 to be similar in magnitude in each regression.

Consistent with the preceding analysis, Table 4 shows that FM data do not convey the same information about student poverty as the other two metrics, which continue to track each other closely. Specifically, the table indicates that a one-percentage-point increase in IPR(130) corresponds to a reduction in student test scores of about 0.017 student standard deviations (noting that IPR(130) indicates the poverty share on a 0-1 scale). The analogous estimate based on the DC share is the same to the thousandth decimal place. In contrast, the same one-percentage-point increase in the FM share corresponds to a much smaller reduction in student achievement—just 0.011 standard deviations. This estimate is about 35 percent smaller in magnitude than the coefficient values from the DC and IPR(130) regressions, reinforcing the finding from above that FM eligibility is oversubscribed and therefore inaccurate as a measure of student poverty.[16]

6.2 Replication with national data

We draw on the Common Core of Data (CCD) to expand our analysis outside of Missouri. First, we construct IPR(130) and IPR(185) estimates for all schools in the U.S. based on the reported IPR values from the NCES. Next, we merge these variables with FM and FRM eligibility shares from the CCD.[17] (We also attempted to replicate our validation regressions of the DC share on IPR(130) in other states, but concluded that the DC data in the CCD are not reliable enough to support the validation regressions.[18])

A challenge with using FM and FRM eligibility shares from the CCD is that some states have changed how they report these categories due to the community eligibility provision (CEP) and others have not, and there is no indicator in the data to distinguish whether reporting is inclusive of CEP coding. It would cloud inference to evaluate a mix of states that are coding their

---

[16] These results are seemingly at odds with related results from student-level regressions in Domina et al. (2018). Domina et al. (2018) run regressions of student test scores on individual poverty status as defined by FM and FRM eligibility, then as defined by IRS data using the same income thresholds, and find larger (more negative) coefficients in the FM and FRM regressions. However, we believe their results have an ambiguous interpretation due to Simpson's paradox as described by Koedel and Parsons (2021). The broader point from their analysis that FM and FRM data contain additional information missed by IRS income data alone is more clearly established, and in results omitted for brevity we confirm this is true in our data as well.

[17] For all 50 states and Washington DC, the CCD includes 105,862 schools in 2017. The SNP metrics from NCES has IPR estimates for 100,191 of these schools—i.e., the coverage rate is 95 percent.

[18] The CCD only recently began including DC data, and the currently-available data are problematic. There are two specific problems. First, the smaller problem is that DC data are missing for most schools in the CCD. Most states report either no DC information or very little DC information. More importantly, the data that are reported in the CCD, at least for MO, are not sufficiently accurate. To test this, we constructed shares of DC, FM, and FRM students in each school using the Missouri microdata and estimated separate univariate regressions of these shares on their data analogs in the CCD, using 2017 data. If the data elements in the MO microdata and the CCD data are the same, we should anticipate coefficients of 1.0 from each of these regressions. For the FM and FRM regressions, our coefficients are close to 1.0, at 0.97 and 0.96, respectively; but for the DC regression, the coefficient is just 0.48. We are not sure what is causing the discrepancy with the DC data, but our MO microdata are surely more reliable because they are based on a direct merge of administrative files between agencies.

data differently. To address this problem, we identify a subset of 27 states in the CCD that do not appear to have manipulated their FRM reporting due to the CEP as of 2017. The criteria we use to identify these states, based on Koedel and Parsons (2021), are (a) less than one percent of schools in the state have a reported FRM share of 100 percent, and (b) there is less than a five-percentage-point increase in the share of schools with missing FRM data between 2014 and 2017. The latter condition reflects the fact that in response to the CEP, some states have begun to report FRM data for schools as missing. The 27 states that satisfy both of these criteria are: AL, AR, CA, CO, CT, FL, HI, IA, ID, IL, IN KS, KY, ME, MI, NC, NH, NJ, NY, OR, RI, TX, VA, VT, WA, WI, and WV.

For each of these states, we run regressions of the FM share on IPR(130), and the FRM share on IPR(185), as shown in equations (5) and (7) above. Like in our analysis of the Missouri microdata, coefficients of 1.0 on the IPR variables would indicate that students' FM and FRM designations are aligned with the stated income requirements of these programs. Coefficients above 1.0 would imply that more students are participating in FM and FRM programs than income-eligibility criteria would suggest. A caveat to our broader analysis using the CCD is that we have not directly validated our IPR(130) measures outside of Missouri. Thus, an additional assumption in this portion of our analysis is that the IPR-based estimates continue to serve as accurate estimates of student poverty shares in other states.

We begin by establishing comparability between the results using the Missouri administrative microdata (from above) and the CCD data. In columns (1) and (2) of Table 5, we show results from regressions of the FM share on IPR(130), and the FRM share on IPR(185), respectively, using the CCD data to populate the FM and FRM variables for Missouri. Note that the Missouri data in the CCD are inclusive of CEP coding, so these results should correspond closely to the results in columns (4) and (8) of Table 2. Table 5 shows that this is indeed the case, confirming that the administrative data and the CCD yield similar results in our analyses of FM and FRM data in Missouri.

Columns (3) and (4) show analogous regressions that pool data from the CCD for all 27 states listed above. We include state fixed effects in the pooled multi-state regressions to isolate within-state variation for identification, although as a practical matter this has no substantive bearing on the findings. The estimates using the 27-state sample are a close match to the Missouri estimates—the coefficients from the FM regressions are very similar, and the coefficient from the FRM regression in the larger sample is slightly lower than the Missouri coefficient. This is broad evidence that FM and FRM data overstate poverty rates. In addition, note that the data for Missouri in the CCD are affected by the CEP, while the 27-state sample was constructed to minimize the influence of the CEP. As a result, the similarity of the estimates in Table 5 from the regressions using FM data suggests that these data overstate poverty more in the 27-state sample than in Missouri.

The pooled multi-state regressions in Table 5 obscure significant state-level heterogeneity in the estimated coefficients on IPR(130) and IPR(185). Figure 2 illustrates this heterogeneity by plotting all 27 state coefficients and their error bands from the FM and FRM regressions. For ease of presentation, states are ordered in each panel from the largest to smallest coefficient values.

The range of estimates shown in Figure 2 is striking. For example, in the FM regressions, the coefficient on IPR(130) ranges from a minimum of 0.50 (Arkansas) to a maximum of 1.75 (Rhode Island). The range of coefficients in the FRM regressions is narrower, but still large, ranging from a minimum of 0.50 (Arkansas) to a maximum of 1.55 (North Carolina). This variability likely reflects a number of factors, including differences across states in districts' FM and FRM certification processes and measurement error in the CCD. The former could reflect, for example, differences in leniency across states in districts' income-verification processes and/or differences in families' willingness to apply to the NSLP. The latter would include all reporting errors between the point of data collection in individual districts to the point of entry into the CCD. In contrast, although we cannot rule out some cross-state variability in the meaning of the IPR(130) and IPR(185) variables, any such variability should be small in comparison to the NSLP data. This is because the data-collection and estimation procedures for these variables are applied uniformly across all states via the NCES estimation procedure (as described in Section 3). The most plausible source of cross-state variability in SNP data is differences in the degree to which public-school enrollment is tied to residential location by state, but the first-order issues driving the cross-state heterogeneity illustrated by Figure 2 are almost surely driven by the FM and FRM data.

In addition to showing that FM and FRM data in the 27-state sample overstate poverty rates similarly to what we find in Missouri, on average, these results also raise larger concerns about using FM and FRM data from the CCD or other sources in multistate studies. Figure 2 suggests these variables have very different meanings with respect to poverty identification across states, implying that using them to control for the same conceptual factors (i.e., poverty or student disadvantage) in research spanning multiple states will be problematic.

Finally, we conduct an analog to the achievement-based analysis shown in Table 4 using the CCD data. Unlike in MO, we do not have access to administrative data on student test scores in the multi-state sample, so we use achievement data from the Stanford Education Data Archive (SEDA) in its place. SEDA contains district-level estimates of average standardized test scores in Math and English Language Arts for students in grades 3-8 throughout the U.S. The comparability across states is facilitated by linking the state tests and the National Assessment of Educational Progress (NAEP) to develop a common scale (Fahle et al., 2018; Fahle, Shear, and Shores, 2019; Reardon, Kalogrides, and Ho, 2021).

SEDA reports achievement at the district level, so we aggregate our poverty data accordingly and estimate regressions of district-level achievement on district-level measures of poverty using the 27-state sample.[19] We construct the district-level poverty shares as enrollment-weighted averages of the school-level poverty shares. We also add state fixed effects to our

[19] SEDA includes district-level average standardized math scores for a national sample of 9,728 out of 10,921 districts in the CCD (89 percent) in 2017. For our selected sample of 27 states, 6,221 out of the 6,853 districts in the CCD (91 percent) have math scores in SEDA in 2017.

regressions, similarly to above, which yields the following analog to equation (8) for this portion of our analysis:

$$Y_{ks} = \varsigma_0 + P_{ks}\varsigma_1 + \nu_s + \eta_{ks} \tag{9}$$

In equation (9), Y_{ks} is the average math achievement level in district *k* in state *s* from SEDA, P_{ks} is the poverty measure of interest, and ν_s is a state fixed effect. In the Missouri-specific version of this model we estimated it three times; defining P_{ks} as the DC share, IPR(130) estimate, and FM share, respectively. For the extended 27-state sample we do not observe the DC share, so we estimate the regression just twice—once defining P_{ks} by the IPR(130) estimate and once defining P_{ks} by the FM share.

The results are shown in Table 6. As in our preceding analysis of the MO data, the coefficient from the regression using IPR(130) is much larger than its analog from the regression using the FM share. Although both coefficients are larger than their comparison coefficients in the Missouri-specific analysis, the relative difference is similar.[20] That is, like in Table 4, the magnitude of ς_1 when we use FM data to measure student poverty is much smaller (by about 53 percent) than when we use IPR(130). This result further supports the conclusion that FM data are not capturing the same level of poverty as IPR(130).

7. Discussion and Conclusion

We validate two external measures of student poverty against each other, then use them to show that FRM data substantially overstate student poverty. This finding is established most clearly in our analysis of the rich administrative data from Missouri. Under the assumption that the validity evidence supporting the accuracy of our IPR(130) estimates in Missouri extends to other states—a plausible assumption but one that we are unable to test directly due to data

[20] The larger absolute values of the coefficients in Table 6 could be driven by a number of factors that differ in this portion of our analysis, including differences in state assessments that could differentially pick up differences in student poverty and disadvantage, factors related to SEDA's process for constructing comparable test scores across states, the impact of district aggregation on the estimates, and the related impact of variability in district size across states that leads to differential aggregation, among other possibilities. As this analysis is only supplementary to our main findings, and the absolute levels of the coefficients are not of first-order importance (it is their relative values that we care about), we did not thoroughly investigate the source(s) of the difference in coefficient magnitudes between our Missouri and 27-state-sample analyses of student test scores.

limitations—this finding is also supported in a larger 27-state sample we construct using CCD data. Community eligibility for free meals within the NSLP has exacerbated the overstatement of poverty in FRM data, but the overstatement precedes the CEP.

Over the course of our investigation of FRM data, we also provide the first evidence of which we are aware on the validity of the school neighborhood poverty metrics recently made available by NCES. Specifically, we show that estimated poverty rates based on the NCES metrics are aligned, on average, with poverty rates calculated using administrative data on direct certification in Missouri. The NCES metrics are publicly available for most schools in the U.S. While our findings do not support the continued use of FRM data to proxy for student poverty, it is of some consolation that we identify a readily-available alternative. It would be useful for future research to continue to push on our initial validity evidence for the NCES metrics by exploring robustness in different locales. Researchers could also consider different manipulations and modeling of the underlying IPR values that we use to construct school-level poverty rates.

We conclude with a brief discussion of why it matters for U.S. education policy that FRM data substantially overstate student poverty. There are two main reasons. First, behaviors of education administrators are a likely source of the oversubscription of students as FRM eligible—e.g., those who are more aggressive in eliciting and approving parental applications can generate higher FRM shares. This should not be surprising based on how the program works (as described above), but it is at odds with the implicit notion that poverty is an objectively-measured attribute. To the extent that FRM-eligible students factor directly into states' funding and accountability systems—which is common—the process by which FRM data are generated, inclusive of potential heterogeneity across states and school districts, raises concerns about behavior that manipulates the underlying data (even if well-intentioned) and the resulting fairness of these systems.

The second reason our findings are important is that they are at odds with traditional uses of historical FRM data, at least prior to the CEP, that at least implicitly assume these data capture poverty levels accurately.[21] A consequence is that pre-CEP FRM rates are being held up as benchmarks by which newer, alternative measures are judged. Or, to be more precise, there are efforts to calibrate newer poverty measures to match older FRM-based measures in schools. The best example of this behavior is the practice of using a multiplier (above 1.0, with a commonly-advocated value of 1.6) to adjust new poverty measures based on direct-certification status to match older FRM-based poverty rates (e.g., see Croninger, Rice, and Checovich, 2015; Grich, 2019). Although accuracy is not the only motivation for this calibration (policy continuity is another motivation), policy documents suggest that it is not commonly understood that the original FRM-based poverty rates are not correct.[22]

In addition to being clear about what FRM data are not, it is also useful to be clear about what they are. While these data do not measure poverty accurately, they are useful measures of student disadvantage, broadly defined. This point is apparent in our regressions of school-average test scores on FRM shares in both Missouri and the larger 27-state sample and reinforced in a more detailed analysis on this point in Domina et al. (2018). In Domina et al., for example, the researchers estimate models of student achievement using both IRS income data and FRM status data and show that both generally contribute to explaining student outcomes. A straightforward explanation—noting that FM and FRM data are oversubscribed relative to their stated income thresholds—is that conditional on true-income based poverty thresholds, these

[21] The idea that the average poverty level is systematically overstated is distinct from (the more apparent) concerns about individual-level FRM designations being coarse and error-prone (e.g., see Harwell and LeBeau, 2010; Michelmore and Dynarski, 2018; Parsons, Koedel, and Tan, 2019).

[22] The original source of the multiplier is the federal legislation that ushered in the CEP. The purpose is to approximate the share of FRM-eligible students in a CEP school based on the share who are directly certified. The latter is the official statistic used to determine CEP-eligibility. The direct-certification rate, times the multiplier, is used by the NSLP to set the meal reimbursement rate under the CEP, with the goal of reimbursing schools at the free-meal rate for all FRM-eligible students. The multiplier is essentially what would be estimated from a regression very similar to the ones we run above of the FM share on DC share, but replacing the FM share with FRM share. The federal guidance establishing the multiplier explicitly references the link to the status quo of using FRM data to measure poverty: "using only the number of directly certified students would result in lower poverty percentages for Community Eligibility schools or LEAs" (U.S. Department of Education, 2014, p. 8). A conceptual reason for the multiplier is that the FRM-eligibility threshold—185 percent of the poverty line—is above the 130-percent level that is effectively enforced via direct certification, so it is intuitive that a multiplier of some sort would be needed to maintain a comparable poverty measure. Implicitly, this seems to be the rationale in the federal guidance, but there is no mention that FRM data overstate school poverty in federal documents or subsequent state documents that discuss similar multipliers (e.g., Croninger, Rice, and Checovich, 2015).

designations identify another tier of disadvantage beyond the tier(s) identified in income data alone (i.e., they identify a group of students who are not as impoverished as students identified by accurate poverty indicators but who are still disadvantaged). It is also possible that the looser rules in assigning FM and FRM eligibility to students permit endogenous behavior by districts, schools, and parents that influence who is approved for the program. For example, locales where students are more disadvantaged along unobserved dimensions may be more aggressive in signing up students for free and reduced-price meals.

One interpretation of this result is as follows: FRM designations are hybrid indicators that combine information about income *per se* and the more nebulous concept of student disadvantage. This shift in the interpretation of FRM data has significant policy implications because it means that older FRM designations were not truly capturing differences in poverty alone, but rather differences in *student disadvantage*. By directly acknowledging this feature of FRM data, it is reasonable to ask whether student disadvantage can be captured more accurately than with FRM data alone. The answer to this question is almost surely yes, because modern education data systems have rich information about students along many dimensions and over time. Examples include, but are not limited to, direct certification (in some states), geographic mobility, attendance patterns, test and other school performance measures, participation in remedial programs, etc. Said another way, when viewed as a measure of student disadvantage, and not poverty *per se*, FRM indicators are ripe for improvement.

This last insight is important because education administrators have been reluctant to make a shift in accountability and funding policies away from what they perceive as poverty measures toward measures of student disadvantage. But this hesitancy is based on the false premise that FRM data have been measuring poverty. While we do not purport to know what the correct decision is with respect to using information about poverty versus disadvantage to inform education policy, establishing a common set of facts is a first step toward productive discussions in this regard.

A final point is that historical FRM data have been used not only at the building level, as we have evaluated them here, but also to track differences between groups of students within schools and districts. The most common application of within-district and -school data has been for the calculation of achievement gaps along the dimension of (presumed) poverty. Even if states and districts wanted to continue to use FRM data for this purpose, the community eligibility provision in the NSLP has made this impossible in many states. Exploring options for identifying high-poverty and/or high-disadvantage students at the individual level is an area of future expansion for this work, but beyond the scope of this article.

References

Bass, D. N. (2010). Fraud in the lunchroom? Federal school-lunch program may not be a reliable measure of poverty. *Education Next*, *10*(1), 67-72. Retrieved from https://www.educationnext.org/fraud-in-the-lunchroom/

Burghardt, J., Silva, T., & Hulsey, L. (2004). Case Study of National School Lunch Program Verification Outcomes in Large Metropolitan School Districts. Nutrition Assistance Program Report Series. Report No. CN-04-AV3. US Department of Agriculture.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633-79.

Chingos, M. M. (2016). No more free lunch for education policymakers and researchers. *Evidence Speaks Reports*, 1(20), 1-4.

Cressie, N. (1989). Geostatistics. *The American Statistician*, 43(4), 197-202.

Cressie, N. (1993). *Statistics for Spatial Data*. New Jersey, NJ: John Wiley & Sons.

Croninger, R. G., Rice, J. K. & Checovich, L. (2015). Evaluation of the Use of Free and Reduced-Price Meal Eligibility as a Proxy for Identifying Economically Disadvantaged Students. Alternative Measures and Recommendations. Denver, CO: Augenblick, Palaich & Associates.

Davis, W., & Musaddiq, T. (2019). Estimating the Effects of Universal Free School Meal Enrollment on Child Health: Evidence from the Community Eligibility Provision in Georgia Schools. Working paper.

de Brey, C., Snyder, T. D., Zhang, A., & Dillow, S. A. (2021). *Digest of Education Statistics 2019.* NCES 2021-009. National Center for Education Statistics.

Domina, T., Pharris-Ciurej, N., Penner, A. M., Penner, E. K., Brummet, Q., Porter, S. R., & Sanabria, T. (2018). Is free and reduced-price lunch a valid measure of educational disadvantage? *Educational Researcher*, *47*(9), 539-555.

Fahle, E. M., Shear, B. R., & Shores, K. A. (2019). Assessment for monitoring of education systems: The US example. *The ANNALS of the American Academy of Political and Social Science*, *683*(1), 58-74.

Fahle, E. M., Shear, B. R., Kalogrides, D., Reardon, S. F., DiSalvo, R., & Ho, A. D. (2018). Stanford Education Data Archive: Technical Documentation (Version 2.1). Retrieved from http://purl.stanford.edu/db586ns4974.

Geverdt, D. (2019). Education Demographic and Geographic Estimates Program (EDGE): School Neighborhood Poverty Estimates, 2016-2017 (NCES 2019-112). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from https://nces.ed.gov/pubsearch/.

Geverdt, D., & Nixon, L. (2018). Sidestepping the Box: Designing a Supplemental Poverty Indicator for School Neighborhoods (NCES 2017-039). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch/.

Gindling, T. H., Mata, C., Kitchin, J., & Avila, E. (2018). Some causes of undercount of low income students under the Community Eligibility Provision in Baltimore City Public Schools (No. 18-01). Department of Economics Working Paper Series Paper.

Gordon, N., & Ruffini, K. (forthcoming). Schoolwide free meals and student discipline: Effects of the community eligibility provision. *Education Finance and Policy*.

Greenberg, E. (2018). New Measures of Student Poverty: Replacing Free and Reduced-Price Lunch Status Based on Household Forms with Direct Certification. Urban Institute.

Greenberg, E., Blagg, K., & Rainer, M. (2019). Measuring Student Poverty: Developing Accurate Counts for School Funding, Accountability, and Research. Research Report. Urban Institute.

Grich, R. (2019). New Strategies for Measuring Poverty in Schools. Explainer. FutureEd. Retrieved from: https://www.future-ed.org/how-states-measure-poverty-in-schools/

Harwell, M., & LeBeau, B. (2010). Student eligibility for a free lunch as an SES measure in education research. *Educational Researcher* 39(2), 120-131.

Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, *26(1)*, 101-136.

Koedel, C., & Parsons, E. (2021). The effect of the Community Eligibility Provision on the ability of free and reduced-price meal data to identify disadvantaged students. *Educational Evaluation and Policy Analysis*, *43(1)*, 3-31.

Lichter, D., Sanders, S., & Johnson, K. (2015). Behind at the starting line: Poverty among Hispanic infants. Durham, NH: University of New Hampshire, Carsey School of Public Policy. Retrieved from https://scholars.unh.edu/cgi/viewcontent.cgi?article=1250&context=carsey

Massachusetts Department of Elementary and Secondary Education. (2017). Low-Income Student Calculation Study. Policy Report from the Massachusetts Department of Elementary and Secondary Education.

Michelmore, K., & Dynarski, S. (2017). The gap within the gap: Using longitudinal data to understand income differences in educational outcomes. *AERA Open*, *3*(1), 1-18.

Moore, Q., Conway, K., Kyler, B., & Gothro, A. (2016). Direct Certification in the National School Lunch Program: State Implementation Progress, School Year 2014-2015 (Report to Congress. Washington, DC: United States Department of Agriculture.

Parsons, E., Koedel, C., & Tan, L. (2019). Accounting for student disadvantage in value-added models. *Journal of Educational and Behavioral Statistics*, *44*(2), 144-179.

Reardon, S. F., Kalogrides, D., & Ho, A. D. (2021). Validation methods for aggregate-level test scale linking: A case study mapping school district test score distributions to a common scale. *Journal of Educational and Behavioral Statistics*, *46*(2), 138-167.

Ruffini, K. (forthcoming). Universal access to free school meals and student achievement: Evidence from the Community Eligibility Provision. *Journal of Human Resources*.

Sandstrom, H., Huerta, S., & Loprest, P. (2014). Understanding the dynamics of disconnection from employment and assistance: Final report. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, US Department of Health and Human Services.

Schwartz, A. E., & Rothbart, M. W. (2020). Let them eat lunch: The impact of universal free meals on student performance. *Journal of Policy Analysis and Management*, *39*(2), 376-410.

U.S. Department of Education (2014). Guidance: The Community Eligibility Provision and Selected Requirements Under Title I, Part A of the Elementary and Secondary Education Act of 1965, As Amended. Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education.

Williams, S. (2013). Public assistance participation among U.S. children in poverty, 2010 (FP-13-02). Working paper, National Center for Family & Marriage.

Zedlewski, S. R., & Martinez-Schiferl, M. (2010). Low-income Hispanic children need both private and public food assistance. Policy Brief, Urban Institute.

Table 1: Summary Statistics, Missouri Data

	2016		2017	
	Mean	Standard Deviation	Mean	Standard Deviation
Demographics				
Black	0.16	0.26	0.16	0.26
Hispanic	0.06	0.08	0.06	0.09
White	0.72	0.28	0.71	0.28
Multi-race	0.03	0.03	0.04	0.03
Asian/Indian/Pacific Islander	0.03	0.03	0.03	0.04
Female	0.49	0.03	0.49	0.03
Minority	0.22	0.27	0.23	0.27
IEP	0.13	0.08	0.14	0.08
ESL	0.04	0.08	0.05	0.09
Test Scores				
Standardized Math Score	0.00	0.45	0.00	0.44
Poverty Measures				
Share Free/Reduced-Price Meal Eligible	0.53	0.26	0.52	0.26
Share Free Meal Eligible	0.47	0.27	0.46	0.27
Share of Directly Certified	0.30	0.18	0.30	0.18
NCES IPR Estimate	284.11	137.88	289.84	138.81
IPR(130)	0.34	0.11	0.33	0.11
IPR(185)	0.42	0.13	0.41	0.13
Avg. Students Per School	423.47	341.14	421.47	340.71
N (Schools)	2,172		2,186	
N (Students)	919,786		921,335	

Notes: This table shows the summary statistics of our analytic sample of schools in Missouri in the 2016 and 2017 school years with at least 25 students in each school in a year. The summary statistics are weighted by enrollment in each school year. Student demographics, test scores, and poverty variables are taken from Missouri administrative microdata. IPR estimates are taken from the NCES school neighborhood poverty (SNP) metrics, and IPR(130) and IPR(185) are calculated from the reported IPR estimates and standard errors for each school. Test scores are from a reduced sample of schools that have test-takers in grades 4-8. The test-taking school samples from 2016 and 2017 include 1,689 and 1,694 schools, respectively.

Table 2: Univariate Alignment Regressions, Missouri Data

VARIABLES	(1) 2016 Dependent variable: DC share	(2) 2017 Dependent variable: DC share	(3) 2016 Dependent variable: FM share	(4) 2017 Dependent variable: FM share	(5) 2016 Dependent variable: FM share	(6) 2017 Dependent variable: FM share	(7) 2016 Dependent variable: FRM share	(8) 2017 Dependent variable: FRM share
IPR(130)	1.026	0.994	1.505***	1.469***				
	(0.033)	(0.034)	(0.048)	(0.051)				
DC share					1.372***	1.386***		
					(0.014)	(0.015)		
IPR(185)							1.385***	1.396***
							(0.034)	(0.037)
Constant	-0.045†††	-0.036†††	-0.046†††	-0.030†	0.049†††	0.050†††	-0.052†††	-0.050†††
	(0.011)	(0.011)	(0.016)	(0.017)	(0.004)	(0.004)	(0.015)	(0.016)
Observations	2,172	2,186	2,172	2,186	2,172	2,186	2,172	2,186
R-squared	0.370	0.348	0.359	0.328	0.849	0.830	0.468	0.440

Notes: This table presents estimates from school level univariate regressions weighted by enrollment in each school year. In each regression, we test the null hypothesis that the poverty-measure coefficient is 1.0; rejection of this null hypothesis at the 1, 5, and 10 percent levels is denoted by ***, **, and *, respectively. For completeness, we also report on the statistical significance of the constant term, where †††, ††, and † indicate the constant is statistically different from zero at the 1, 5, and 10 percent levels, respectively. Standard errors are in parentheses.

Table 3: Univariate Alignment Regressions with Imputed FM and FRM Shares for CEP Schools, Missouri Data

VARIABLES	(1) 2016 Dependent variable: FM share	(2) 2017 Dependent variable: FM share	(3) 2016 Dependent variable: FRM share	(4) 2017 Dependent variable: FRM share
IPR(130)	1.354***	1.318***		
	(0.040)	(0.043)		
IPR(185)			1.293***	1.308***
			(0.030)	(0.032)
Constant	-0.020	-0.008	-0.033††	-0.035††
	(0.014)	(0.014)	(0.013)	(0.014)
Observations	2,160	2,168	2,160	2,168
R-squared	0.398	0.373	0.504	0.490

Notes: This table presents estimates from school level univariate regressions weighted by enrollment in each school year. For CEP schools, the FM and FRM shares are imputed at the 2014 level, the last year of non-CEP coded data in Missouri. In each regression, we test the null hypothesis that the poverty-measure coefficient is 1.0; rejection of this null hypothesis at the 1, 5, and 10 percent levels is denoted by ***, **, and *, respectively. For completeness, we also report on the statistical significance of the constant term, where †††, ††, and † indicate the constant is statistically different from zero at the 1, 5, and 10 percent levels, respectively. Standard errors are in parentheses.

Table 4: Relationships between Student Test Scores and Measured Poverty, Missouri Data

VARIABLES	(1) 2017 Dependent Variable: School Avg. Test Score	(2) 2017 Dependent Variable: School Avg. Test Score	(3) 2017 Dependent Variable: School Avg. Test Score
IPR(130)	-1.732***		
	(0.167)		
DC share		-1.685***	
		(0.056)	
FM share			-1.087***
			(0.043)
Constant	0.577†††	0.535†††	0.527†††
	(0.058)	(0.020)	(0.021)
Observations	1,694	1,694	1,694
R-squared	0.172	0.490	0.454

Notes: This table presents estimates from school-level univariate regressions where the dependent variable is the school average standardized math test score, and the independent variables are three different measures of poverty–IPR(130), the DC share, and the FM share in the school. All regressions are weighted by enrollment. In each regression, we test the null hypothesis that the poverty-measure coefficient is zero; rejection of this null hypothesis at the 1, 5, and 10 percent levels is denoted by ***, **, and *, respectively. For presentational consistency, we continue to denote statistical significance of the constant term at the 1, 5, and 10 percent levels using the same †††, ††, and † indicators from previous tables. Standard errors are in parentheses.

Table 5: Univariate Alignment Regressions, Missouri and the 27-State Extended Sample Using the Common Core of Data

VARIABLES	(1) 2017 Dependent Variable: School FM Share in MO	(2) 2017 Dependent Variable: School FRM Share in MO	(3) 2017 Dependent Variable: School FM Share in 27 States, Not CEP-Coded	(4) 2017 Dependent Variable: School FRM Share in 27 States, Not CEP-Coded
IPR(130)	1.419***		1.397***	
	(0.052)		(0.008)	
IPR(185)		1.363***		1.253***
		(0.040)		(0.007)
Constant	-0.030†	-0.050†††	-0.045†††	-0.025†††
	(0.017)	(0.017)	(0.006)	(0.006)
State Fixed Effects	N/A	N/A	Yes	Yes
Observations	2,257	2,257	61,270	61,270
R-squared	0.306	0.417	0.477	0.525

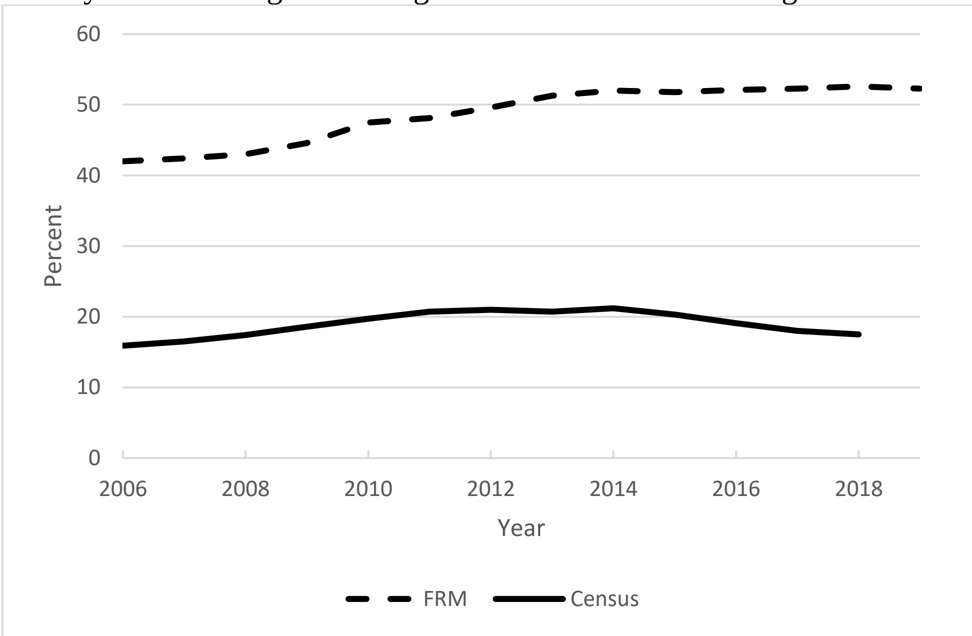
Notes: This table presents estimates from school-level univariate regressions weighted by enrollment in each school in 2017 using CCD and SNP data from NCES. In each regression, we test the null hypothesis that the poverty-measure coefficient is 1.0; rejection of this null hypothesis at the 1, 5, and 10 percent levels is denoted by ***, **, and *, respectively. For completeness, we also report on the statistical significance of the constant term, where †††, ††, and † indicate the constant is statistically different from zero at the 1, 5, and 10 percent levels, respectively. Standard errors are in parentheses.

Table 6: Relationships between Student Test Scores and Measured Poverty, 27-State Extended Sample Using the Common Core of Data and Stanford Education Data Archive

VARIABLES	(1) 2017 Dependent Variable: District Avg. Test Score	(2) 2017 Dependent Variable: District Avg. Test Score
District IPR(130)	-3.020***	
	(0.081)	
District FM Share		-1.404***
		(0.047)
Constant	0.815†††	0.364†††
	(0.043)	(0.036)
State Fixed Effects	Yes	Yes
Observations	6,221	6,221
R-squared	0.579	0.692

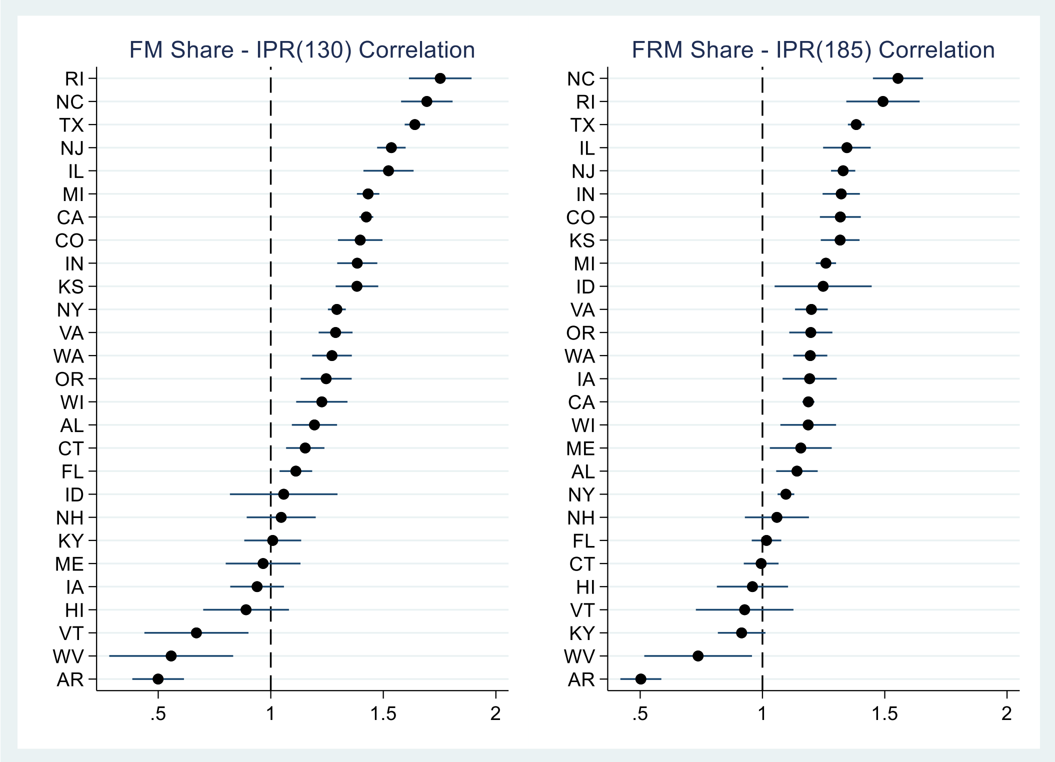
Notes: This is a national-level analog to Table 4 using data from CCD, SEDA, and SNP data from NCES. This table presents estimates from district-level univariate regressions where the dependent variable is the district average standardized test score, and the independent variables are IPR(130) and FM share in the district, respectively. The regressions are weighted by enrollment in each district. In each regression, we test the null hypothesis that the poverty-measure coefficient is zero; rejection of this null hypothesis at the 1, 5, and 10 percent levels is denoted by ***, **, and *, respectively. For presentational consistency, we continue to denote statistical significance of the constant term at the 1, 5, and 10 percent levels using the same †††, ††, and † indicators from previous tables. Standard errors are in parentheses.

Figure 1: Poverty Rates Among School-Aged Children Measured Using Different Data Sources



Notes: Trends in the share of FRM students and the share of school aged children living in families with incomes at or below the poverty line. Source: Digest of Education Statistics.

Figure 2: Heterogeneity of the FM and FRM Regression Coefficients in the 27-State Sample



Notes: The left panel shows the estimated coefficients from univariate regressions of the FM share on IPR(130) for the 27 states with non-CEP coded data, along with 95 percent confidence intervals. The right panel shows analogous coefficients and confidence intervals from univariate regressions of the FRM share on IPR(185). States are ordered in descending order of the coefficient values in each panel. The data sources are the CCD and SNP data from NCES.

Appendix A
Correspondence between the DC share and IPR(130) throughout the distribution

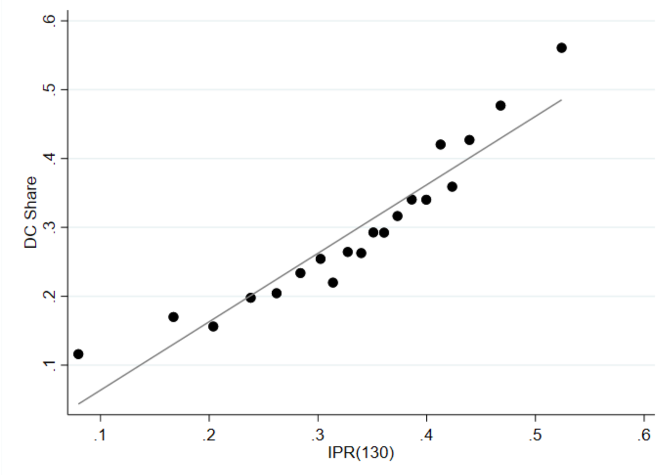
The analysis in the main text shows that IPR(130) and the DC share are validated against each other *on average* in the Missouri data. This is sufficient to support our investigation of FRM data, which is also conducted on average. In this appendix, we document the relationship between IPR(130) and the DC share throughout the distribution, with the goal of informing future uses of IPR(130) (and other IPR(X) metrics more broadly).

Figure A1 provides a binned scatterplot of IPR(130) and the DC share using 2017 data. Theoretically, if IPR(130) and the DC share are measuring identical constructs and contain no error, we would expect a linear plot on the 45-degree line. Instead, the data plot is slightly convex—the relationship between the two variables is flatter than the 45-degree line at low values of IPR(130) and steeper than the 45-degree line at high values of IPR(130). Because both metrics are estimates of poverty and subject to limitations, and we do not observe "true" poverty values, it is difficult to identify the source of the modest nonlinearity in the figure. One explanation is that there are subpopulations of students who are underrepresented relative to their poverty shares in the programs that lead to direct certification and these students are not evenly distributed in the income distribution. Previous research suggests that Hispanic students comprise one such group (Lichter, Sanders, and Johnson, 2015; Sandstrom, Huerta, and Loprest, 2014; Williams, 2013; Zedlewski and Martinez-Schiferl, 2010), and there may be other groups that are harder to identify.[23] It may also be that there is modest heterogeneity in the efficacy of the Kriging procedure used by NCES to estimate average incomes for individual schools or in the efficacy of our conversion process to identify students at or below 130 percent of the poverty line.

It would be convenient if the relationship in Figure A1 were perfectly linear, but noting that it is not, our hope is that future research can shed additional light on the strengths and weaknesses of these measures and perhaps improve on them. We are optimistic that these measures, or similar measures, can be useful diagnostic tools for researchers and policymakers interested in measuring student poverty.

[23] The demographics of Missouri are such that the impact of Hispanic underrepresentation in programs that lead to direct certification will be limited (see Table 1).

Appendix Figure A1: IPR(130)-DC Share Binned Scatterplot (2017)



Notes: We constructed this chart by dividing IPR(130) into 20 equal-sized bins—each dot indicates the mean values of IPR(130) and the DC share within each bin. The full ranges of the IPR(130) and DC-share variables are 0-0.72 and 0.01-0.91, respectively.

Appendix B
Supplementary Tables

Appendix Table B1: Mean Squared Error Analysis

	2016	2017
FM share	0.051	0.054
IPR(130)	0.021	0.022
Observations	2,172	2,186

Notes: This table shows mean squared errors (MSEs) for the FM share and IPR(130) values for schools in the Missouri administrative microdata in 2016 and 2017. These MSE calculations assume the share of DC students reflects the true poverty share, which is almost surely incorrect but is likely approximately accurate. Smaller values indicate less error.