Issues in Rural Education: A Review of the Literature and New Evidence from Administrative Data.

Elton Mykerezi
Department of Applied Economics
Extension Center for Community Vitality
University of Minnesota, Twin Cities

Abstract:

This report provides an overview of some persisting issues in rural education as well as a discussion of emerging topics, largely from an economic perspective. The author discusses challenges that education institutions face in a rural context, as well as ways in which problems in rural schools are similar to those in their urban counterparts. The report shows how many of the problems that often plague rural school districts, such as difficulties with recruitment and retention of effective teachers, are fundamentally problems that stem from the way that teacher markets are organized nationwide. Some conditions that are unique to rural areas may exacerbate difficulties that exist in most teacher markets (e.g. sparse populations and thin markets), but they also have the potential to ameliorate the situation (e.g. a smaller wage premium for college graduates in the rural private sector).

Several policies and practices that have either been widely discussed in the literature as potential solutions to problems in rural education or that show particular promise for resolving problems unique to rural schools are then discussed critically. Examples include some well-established strategies, such as pooled provision of certain services and/or school consolidations across districts, distance learning, several teacher recruitment/retention initiatives, special allowances in education funding formulas and four day school weeks, among others.

Advances in Economics of Education research and recent educational policies and practices implemented in various diverse educational settings nationwide are then discussed in the context of potential applicability to rural schools. These include teacher performance measurement, selection and retention, professional development and pay for performance contracts, among others. The author concludes by distilling Implications for rural education policy and suggesting areas of research that are likely to prove productive for future inquiry on rural education.

1. Background

Human capital is a key prerequisite to economic success and it has historically been relatively scarce in non-metropolitan U.S. communities. Despite major improvements over time, a sizable rural-urban gap in education among adults has continued to persist over the last four decades, albeit with divergent trends across different levels of education. While the metro-non-metro gap in adults with no high school degree has been closing, the gap in adult college graduates has widened. Figure 1.a shows trends in shares of adults without a high school degree,

while Figure 1.b shows adult shares with at least a BS degree. Decennial census data (up until 2000) and five-year aggregates of data from the American Community Survey (for 2006-10 and 2008-12)[1] show that the incidence of adults with no high school degree has declined substantially across the board since 1970. The rural urban gap in 1970 was a full 10 percentage points, and it has declined to 2.7 percentage points. The incidence of adults with a college degree has also increased across the board, but growth in metro areas has outpaced that in non-metro areas, increasing the gap from less than 5 percentage points to nearly 13 in the 2008-12 period.

The persistence of such gaps in education among adults, combined with the fact that population sparsity can create unique obstacles for schools, has led some to question the performance of non-metro schools. Educational gaps across geographic areas are generally hypothesized to arise due to the combined effect of three sets of circumstances, socioeconomic background, migration and educational institutions (Mykerezi et al., 2014). The geographic distribution of demographic and socioeconomic characteristics is uneven across urban and rural counties; many rural counties are persistently poor, while others have high concentration of low income minorities in certain regions (e.g. Mykerezi and Mills, 2004). Migration patterns across rural and urban counties also differ by level of education, with those who have college degrees being far more likely to move than those without (e.g. Mills and Hazarika, 2001). Finally, factors uniquely related to the non-metro school settings could play a role in educational gaps.

Naturally, many studies have proposed hypotheses or introduced more comprehensive constructs (e.g. Roscigno and Crowley, 2001) to help understand the potentially unique aspects of education in rural areas from a theoretical perspective. Many other authors, on the other hand, have attempted to disentangle the influence of demography and socioeconomic background from institutional factors related to education in non-metro areas, at educational outcomes in different stages of the educational process; examples include achievement before high school (e.g. Durham and Smith 2006), during high school (e.g. Reeves 2012; Fan and Chen 1999), gaps in high school dropout rates (e.g. Jordan, Kostandini and Mykerezi, 2012) and college completion (e.g. Gibbs, 1998; Byun, Meece, and Irvin 2012).

2. Do Rural and Urban Schools Produce Different Outcomes?

2.1 Theoretical Perspective on Rural-Urban Differences

The scope of this section is to focus on challenges that rural schools face and potential solutions. I first discuss conditions under which rural-urban differences in educational outcomes would emerge from a theoretical perspective, then provide a brief review of the empirical literature on observed rural-urban differences.

[1] Data from the American Community Survey can be disclosed at the county level for all US counties only when they are aggregated over several years, in order to protect the confidentiality of survey respondents. Prior to the release of the ACS, nationally representative data with a broad coverage of topics were only available for all U.S. counties from the decennial census survey.

Population density is viewed as the fundamental feature distinguishing rural and urban areas, both in theoretical representations and in how definitions of "rurality" are operationalized for empirical purposes (e.g. OMB definitions of "metro" and "non-metro" areas). Definitions of "rurality" then use presence of areas that are sufficiently dense to be considered as "metro" and distance from such concentrations for categorizing meaningful political units, such as districts or counties, as rural. Thus, low density, higher geographic isolation and smaller size are features that are uniquely associated with rural counties or school districts (often the units of empirical analysis). As a result, the industrial composition and provision or public services is likely one that does not take advantage of economies of size, but relies on availability of space, land and natural resources. This can lead to lower returns to human capital, lower levels of human-made amenities (e.g. restaurants, professional sports, medical services, transportation options, etc.).

Of course, none of these features are ubiquitous in rural, or absent from urban areas (except for the low density, by definition). Other features that need not be, but are often disproportionately found in rural areas include high and persistent poverty, low education among adults and older populations (e.g. Kusmin, 2013). It must be noted that there are substantial differences within what is typically defined as rural, as well as regional and state differences that are intentionally over-simplified in this discussion, since most literature on rural-urban differences is targeted at capturing differences that apply to a broad range of rural areas.

To illustrate how these identifying features of rural areas impact education, I start by representing school operations via a simple static production function, whereby inputs are used to generate some educational outcome, as in Andrews, Duncombe and Yinger (2002). In this setting, schools use purchased inputs (x) to produce a range of school activities (G) according to $G=f(x)$. The educational outcome of importance to parents, policymakers and taxpayers, denoted by "S" (typically operationalized by some student outcome, such as test scores or graduation rates) depends on school activities ($G(x)$) as well as student, family and neighborhood characteristics (E), enrollment size of the school and district (N), and other unobservable district or school specific effects (d). The production function for student performance is then represented as $S= h(G, E, N, d)$ or $S=h(f(X), E, N, d)$.

In this framework, rural and urban schools can differ in terms of quantities of inputs available (x) per pupil (N), as well as family and neighborhood socio economic characteristics (E), effectively allowing empirical differences in observed student outcomes to be decomposed into those due to background and resources and "other factors". Because rural schools are often small, returns to enrollment are often of interest, and they can captured by the elasticity, $(dS/dN)(N/S)$. Further, the term (d) can be allowed to capture other systematic differences, such as social capital (e.g. Israel et al., 2001).

This representation is important because it allows analysis related to the central feature of rural school districts (sparsity) to be analyzed in a manner that is consistent with the literature on education production and finance in general. Schools only receive a small amount of financing from the federal government, and are typically financed via a state grant that depends on the number of students served (often adjusted by type of student, in terms of age, special needs,

poverty status, and other factors that states may deem to affect the marginal cost of educating each type of student). The rest of the funding is provided by local governments, most often using local property tax revenue. For rural districts, sparsity can imply very low revenue levels because of low enrollment and, possibly lower local tax base. Popular strategies to cope have included service sharing across districts and school consolidations to take advantage of scale at the district level, borrowing or taxing extra resources locally (through operating or capital revenue bonds and/or levy referenda).

Special adjustments of the funding formulae at the state-level to account for sparsity is also common; more than half of the states have some adjustment in place to provide extra revenue to sparsely populated districts. Both, the problems and solutions then amount to identifying optimal school sizes, optimal roles/levels of local governance (in terms of what is provided locally and what is shared at higher levels) and local willingness to pay for schools. This allows rural districts to learn from broader research on education and also suggest strategies for coping. For instance, using the school as a civic/community center can accomplish some cost sharing and it can increase the taxpayer's marginal willingness to pay for the schools. A literature review of school financing and strategies for rural areas will follow in the next chapter.

However useful, this simple model makes two crucial assumptions that restrict its usefulness for analyzing some important issues in education. First, agents who optimize a production activity are typically able to purchase any level of input X that they choose to at some market price. This is not the case in education; the price of the most important input, teachers, can be quite inflexible, causing teacher labor markets not to clear, and generating excess supply of some types of teachers and excess demand of other types of teachers. This causes districts to find innovative ways to compensate teachers that are in short supply and teachers who are in demand often seek non-pecuniary rewards, such as a desirable teaching environment. This causes substantial recruiting issues for many different kinds of schools, including many in rural areas. A fuller discussion of teacher labor market dynamics is deferred to chapter 3.

The second issue with the static production function approach is that it does not allow for dynamic feedback effects in resources and educational investments. Much of the empirical work on rural-urban educational achievement gap in rural sociology is based on a framework proposed by Roscigno and Crowley (2001), which distinguished between "resources" and "investments" in the public and the private sector. This construct was then used as a basis for estimation of rural-urban gaps in test scores for 8^{th} graders, and has been used widely in cross-sectional analysis of rural-urban achievement gaps.

Effectively, their theoretical construct implies a dynamic two-generation household model with endogenous investments in the public sector (school spending x that depends on endowments and expected returns to output S), where households self-select into public education bundles. These generally take a 'micro' approach, focusing on how individual households make decisions related to location, participation in public life and educational investments. Such theories postulate that individuals (and their families acting as proxies early in life) make location and educational decisions aimed at improving the long-term wellbeing of

youth. Households chose locations with tax structures and public good provisions that are consistent with their long-term aspirations (e.g. Tiebout 1956), then they factor existing private and public resources and access to credit markets to make educational investments that maximize the youth's lifetime wellbeing (Becker 1964). This self-selection mechanism alone highlights the possibility that not only are resources clustered geographically, but so are intentions to invest out of any given resource base (because of the purposeful self-selection into the desired public education menu). Youths themselves, are influenced by this environment and make education, labor force participation and migration decisions (Becker, 1964; Sjaastad 1962; Mills and Hazarika 2001) which will, in turn, continue to affect the community's labor markets and resource base.

Further, as noted, returns to education are lower in rural areas. Educated youth can chose to migrate after schooling, and do so disproportionally from rural areas (Mills and Hazarika, 2001). Human capital theory proposes that would-be migrants incur migration costs (both monetary and psychological), and a decision to move is made if these costs are outweighed by lifetime benefits from migration, not the least of which are realized as differential returns to skills in the labor market (e.g. Mills and Hazarika 2001; De Jong and Fawcett 1981; Bowles 1970). The rural-urban earnings gap by education level, combined with the proposition that migration has inherent costs, also has implications for educational incentives among incumbent nonmetro youth. As noted, educational investments are made as a matter of balance between (discounted) lifetime costs and benefits from education (Becker 1964). Nonmetro youth can only access the high-return market in metro areas for skilled occupations if they are willing to incur the costs of migration, thus, the net ex-ante return to schooling is lower for nonmetro than for metro youth. Migration may also have a feedback effect in public incentives to invest in education; nonmetro areas may invest less in public education, since they are less likely to reap the long term benefits of such investments (Swaim and Teixeira 1991; Roscigno and Crowley 2001).

In sum, this is a dynamic system fueled by sequential decisions over multiple generations as well as multiple life stages within each generation. Families with different resources and educational aspirations self-select into locations, defining resources available and preferences for educational investments. Lower wages in general and lower returns to education in non-metro areas then supply asymmetric incentives to invest in education and to stay in non-metro areas for those who do invest. The potential for lower investments in education and higher out-migration of educated adults, in turn, perpetuates lower wages and resources in non-metro areas.

2.2. Empirical Evidence on Rural-Urban achievement gaps

Most of the literature on achievement gaps uses cross-sectional or short panel data and various regression based methods to identify the share of any rural-urban achievement gap that can be traced to resources and investments, and treat the remainder as an unexplained rural effect (often allowed to vary with family and community characteristics). Overall, results have been

mixed. A review of the literature from the 1990s, Khattri, Riley, and Kane (1997) found no consensus on weather rural schools produce a disadvantage. More recent studies have also produced little consensus. The primary shortcoming of recent studies is that they are neither able to estimate full structural models with cross-sectional data, nor do they take a time-indexed approach that takes account of the fact that learning is a dynamic process, even when they have access to longitudinal data (see Todd and Wolpin, 2001). For instance, Roscigno and Crowley's (2001) finding that rural-urban disadvantages in test scores in adolescence can be explained by variables that can be thought of as resources and investments at the household and school level does not imply that investing in high schools can close rural-urban achievement gaps. Because there are no controls for lagged achievement or lagged resources and investments, it is possible that achievement gaps started well before high school, and can be remediated with early investments. The fact that measures of investment in tenth grade help close observed gaps could just reflect serial correlation in investments.

Mykerezi et al., 2014 provides a recent review of the literature and takes a reduced form approach with long-term longitudinal data to track possible rural-urban achievement gaps at various ages of the same cohort. The study tracked achievement gaps of NLSY 97 respondents from age 12 to 30. First, rural urban gaps for those who were 13 years of age or younger were estimated (random third of the sample), then for those aged 16 or older, to examine if any gaps open before high school, and if they widen or narrow during. Then graduation rates are examined (controlling only for prior achievement), as are the rates at which college preparatory tests are taken and performance on those tests (ACT/SAT), then college attendance (controlling for prior achievement and intent, as proxied by taking high stakes college tests). Finally, the impact of prior achievement and attainment on migration and wages was examined. The study found that rural-urban gaps in cognitive test scores emerge before high school, then remain constant through high school. Rural and urban youth were also found to graduate high school and take a college admission test at the same rate. Performance on these tests was only found to be lower to the extent that would be predicted by the pre-existing achievement gap measured at 8th grade.

Non-metro youth also obtained college degrees at a lower rate, despite the similar rates of college prep test taking. Only, half the disadvantage in degree attainment was explained by prior cognitive achievement gaps. A substantial gap in college attendance remains, even after controlling for background, college admission tests and other measures of prior achievement.

2.3. Are Rural Tax-Payers Willing to Pay Less? Evidence from Referendum Data in Minnesota. (Incomplete)

The fact that no major differences in in outcomes are consistently found (except for in early schooling) may imply that some of the theoretical advantages of living in rural areas (e.g. social capital) may, on average, outweigh some of the negatives. Alternatively, noisy measures of rurality, student outcomes, etc., may conceal smaller effects of conditions in rural areas on achievement. One avenue of future research is to pursue research that is directly driven by

theory, rather than attempt to capture "the rural effect" as a residual. As an example, referendum data on all districts from 1992 until 2013 in Minnesota are used in this section to examine if rural districts tend to turn to their voters less frequently, and if percent voting yes in rural areas (and probability of a win) is systematically different, conditional on putting a levy on the ballot in the first place.

(Coming up).

3. Selecting and Retaining Effective Teachers:

3.1 Pay by Occupation and Degree (ACS)

From a theoretical perspective, teachers choose to start and switch careers by considering the net benefit from competing occupations (e.g. Dolton and Van der Klaauw, 1999)

Specifically, teachers are almost explicitly paid based on their highest degree and years of experience within each district. Field of study plays little role in determining pay.

To understand why this wage compression leads to teacher "shortages" in some disciplines and a "surplus" in others it is useful to consider a national profile of earnings by college major and occupation. Appendix Table 3.2 presents nationally representative data on "lifetime earnings profiles" in millions of dollars based on US Census Bureau's American Community Survey (a survey of approximately 50,000 households each month). Majors are presented left to right and occupations top to bottom. For instance, the first cell indicates that the average college graduate of any major and any occupation will make $2.4M over an average career.

Figure 2 extracts earnings by major for all college educated workers, those in management, and those in education. The first thing to notice is that education is a low paying occupation, on average. It pays less than the average for all majors, regardless of major. So attracting individuals in education requires that they accept lower paying jobs, on average, than they would get in the private sector.

To understand the problem of shortages, in Figure 2 we compare the earnings profiles by major of education as an occupation (also in row 11 in appendix Table 1), with that of management (also in row 4 in appendix Table 1), and the average college educated worker (also in row 1 in appendix Table 1). Management provides a good illustration because it is another occupation that requires a broad range of preparations, backgrounds and skills. Regardless of college major, those working in education will make between $1.8M and $1.9M over their careers. In management, on the other hand, earnings will vary substantially depending on college major. Earnings vary from $4.1M for engineering majors to $2.7M for psychology majors, to $2.3M for education majors working in management. This explains why there are more "shortages" and "surpluses" in the market for teachers than that for managers. The market for managers adjusts

wages to attract the skills it needs, paying nearly twice as much for skills that are in short supply, but the educational market tries to hire all skill sets at the same price.

It is, thus, no coincidence that STEM appears on the "teacher shortage" list while social studies, art and literature are chronically over supplied. The average individual with a BS degree in engineering or math would have to give up between $3.1M and $3.3M, on average, over their careers to make $1.8M-$1.9M in education. The average person with an arts degree would only make $2M, on average. So, at a price of $1.8M-$1.9M it is far easier to attract individuals with a background in arts and humanities, whose opportunity cost is only $2M than those trained in disciplines that require quantitative skill, whose opportunity cost is over $3M.

Further, a look at education majors (last column in the table) shows that those who have the skill set and chose to take on occupations like Math, Engineering and Management do make substantially more than they would in education. So, all in all, it is harder to pull individuals with quantitative majors into education and easier to lose education majors with a quantitative specialization to other occupations.

One final point on wage disparities: the primary tool used to attract teachers in chronic shortage areas, loan forgiveness, is not likely to be very successful. Average student debt ranges between $15-$30K total while differences in lifetime earnings can easily be in the hundreds of thousands, or even millions.

3.2. Recruiting Teachers in Rural Areas (incomplete)

Many authors point out that Rural schools may find it particularly difficult to recruit and retain effective teachers. For instance, Cowen et al., (2012) show that teachers are much more likely to leave a remote area (Appalachian Kentucky, in this case) than to move in to one.

There is also evidence that rural teachers are less likely to have a graduate degree, more likely to have a BS in teaching (as opposed to the subject matter, such as Math), less likely to have completed licensure, but no consistent evidence that they exit the profession at a younger age than urban teachers.

However, these signals do not have much predictive power over teacher value added. Evidence that rural schools may find it more difficult to hire in certain areas does not translate into evidence that they hire lower value-added teachers in general. Research is needed to examine rural-urban differences in states that have now implemented universal evaluation systems.

3.3 Teacher Exits by Subject and Rurality: Evidence from Minnesota (Incomplete)

We gain some perspective on demand by subject area and rural location by examining exit patterns of the population of Minnesota teachers for school years 2009-2010 until 2012-2013. The next two graphs plot the experience of teachers who exit between 2010 and 2012, at the time at which they exit.

Specifically, Figure 3.5 shows that for teachers who taught arts, language, literature, humanities and social sciences, experience when they exit the profession resembles quite closely patterns that are typical for all other teachers in the state averaged together. Some testing across multiple categories showed that the two most important differences can be seen among the exit patterns of general education elementary school teachers and those licensed to teach math, sciences or technology[2]. Figure 3.6 plots experience at exit for general education elementary school teachers and STEM teachers regardless of grade. Notice the much higher density of young teachers exiting in the first 10 years for STEM than for elementary education teachers. The opposite is true for those exiting after having taught for 25 or more years.

3.5. Importance of Predicting Teacher Effectiveness: A National Perspective

Teacher effectiveness has a substantial impact on a student's success in school and, eventually, in the labor market. Over the last two decades, a huge amount of scholarly research has focused on teacher quality; measuring it, examining its predictors, consequences for student outcomes and responses to policy.[3] Estimates suggest that a teacher who performs at one standard deviation above the mean effectiveness generates over $400,000 in net present value each year (Hanushek, 2011; Chetty et al., 2011).

Not surprisingly, increasing teacher effectiveness is one of the main preoccupations of education policy makers at all levels of responsibility around the country. The proposed project tackles teacher effectiveness measurement at MPS and pre-hire predictors of teacher effectives. In doing so it will generate actionable knowledge for MPS, valuable information for practitioners in general, and novel scholarly research.

Research to-date has established a few stylized facts about teachers. First, there is large and persistent variation in teacher effectiveness[4], even within schools. Second, variation in teacher

---

[2] Other areas, such as foreign languages also showed particularly early exit patterns, but the number of teachers in that area is typically a small sample, not warranting separate treatment. Other areas, which are large and show shortages in administrator reports, such as Special Education, were not found to have exit patterns that were particularly different than all other teachers. Thus, we focus attention on general education elementary (in chronic over-supply according to MN school administrators) and STEM (in chronic shortage), because they comprise a large share of teachers and have sufficiently different exit patterns.

[3] See Jackson, Rockoff & Staiger (2014) for a recent review the literature.

[4] Effectiveness is most commonly measured by value added measures, although not explicitly. Measurement of teacher effectiveness is subject to some debate. In this section I assume we are able to measure effectiveness in an unbiased and consistent way. In the next session I discuss measures of effectiveness in the context of the current state of the academic literature.

effectiveness is not predicted by factors that are typically used for most personnel decisions (education, licensure, etc.). Neither these simple signals nor decision-makers' discretion are successfully differentiating between effective and ineffective teachers. Schools could selectively employ teachers, either through improvements in recruiting and screening or through selective retention. Unfortunately, researchers have yet to identify reliable predictors of teacher effectiveness observable at the time of hire (Clotfelter, Ladd & Vigdor, 2010; Hanushek & Rivkin, 2012; Harris & Sass, 2011).

Absent clear signals of effectiveness at the time of hire, schools must hire based on very noisy signals and then observe the teacher during a probationary period. In theory, the system is set up such that if during the probationary period the teacher proves to be effective, she is granted tenure and if she is ineffective, she is dismissed. However, this on-the-job monitoring is expensive because many children are subjected to what turns out to be ineffective teaching. Additionally, a long-run equilibrium with wide application of performance based dismissals may make the profession less desirable and either affect the quality of the pool of new entrants or demand higher wages (e.g. Rothstein, 2012).

Performance-based retention also faces significant political and programmatic barriers, especially in settings where collective bargaining plays a large role (West & Mykerezi, 2011). For instance, stimulations show that replacing the most ineffective 5% of current teachers with teachers of average effectiveness would generate about $10k annually per classroom in net present value (e.g. Hanushek, 2011; Goldhaber & Hansen, 2013; Chetty et al., 2013). Yet, in practice, almost all teachers are granted tenure. Separations have been found to be largely determined by seniority and uncorrelated with measured teacher effectiveness (Boyd et al., 2011; Goldhaber & Theobald, 2013).

While any significant policy of performance based dismissals is yet to appear in practice, research on the optimal number of years for the probationary period and the optimal number of teachers granted tenure suggests that, absent information about effectiveness at the time of hire, optimal dismissal rates may need to be in excess of 80%. If pre-hire signals with some predictive power were identified, the optimal proportion of dismissals falls (Staiger & Rockoff, 2010).

Improvements in ability to predict effectiveness during the hiring process, thus, could generate feasible policies with large social benefits. The primary obstacle to such policies is an absence of meaningful predictors of teacher quality observable before hire. Most commonly observed teacher characteristics such as education, degree granting institution, licensure, etc., are either uncorrelated with teacher effectiveness or explain little of the variation in effectiveness across teachers (Jackson et al., 2014).[5]

[5] Experience is an exception, in that it has been shown to predict effectiveness for early-career teachers (e.g. Rockoff, 2004; Rivkin et al., 2005), and latter (Wiswall, 2013). However, there is still substantial variation *within* seniority levels.

Two recent developments, however, show that active and purposeful efforts to collect more data before hire could generate significant predictors of teacher effectiveness. Rockoff et al. (2011) conducted a survey of new math teachers in New York City to collect data on characteristics linked to performance in other professions, such as 'the big five' (a hierarchical inventory of personality traits that is thought to capture most variation in personality traits across people). No single variable predicted value added, but roughly 10 percent of the variation in new teacher performance could be predicted using indices that combine the characteristics. This study, however, did not use data that were collected as part of an actual teacher selection process.

The only study that examined how data that were collected as part of an actual selection process performed was Dobbie (2011). The study used data from Teach for America (TFA), where candidates are rated on several criteria prior to hire and found that prior academic achievement (college GPA in the last two years), leadership (scored based on activities reported in resumes), and perseverance (elicited via a survey of past experiences) predict student gains in math in a teacher's first year. Even though TFA is rather selective, it may differ from typical teacher recruiting settings in important ways.

TFA has typically targeted urban areas, but as of 2012, they started a rural school initiative. Rural districts could seek partnerships with TFA. Not only would they help staff hard to staff position, but TFA is the only organization thus far that has been able to predict teacher performance with tools intended for high-stakes job interviews (perhaps not too high stakes for tfa?)

3.3 Measuring Teacher Quality

Attempts to measure teacher effectiveness have spurred a substantial amount of discussion. Though VAM that rely on standardized test score growth are widely computed and used for various purposes, their merits have been subject to vigorous scholarly debate (e.g. Hanushek, & Rivkin, 2012). VAM are appealing because they are output-based and readily computable with longitudinal, databases. Also, recent research has shown that VAM predict desirable future outcomes. Being assigned a high value-added teacher in grades 4 to 8 was found to be positively associated with the odds of attending college, adult salaries, higher savings rates, and lower teen pregnancy rates (among female students) (Chetty et al., 2013).

However, concerns have been raised about possible bias in VAM and low persistence over time and setting (making teacher-quality measures context specific) (e.g. Jacob, Lefgren, & Sims, 2010; Rothstein, 2010). Use of VAM in high stakes personnel decisions has also been criticized because it can create incentives to game by 'teaching to the test' or even flat out cheating (e.g. Jacob & Levitt., 2003). Finally, VAMs are only applicable to the subset of licensed

teachers who are either solely or mostly responsible for teaching cohorts of children in tested grades and subjects.

Teacher performance is complex and multidimensional, thus districts are increasingly using multiple, complementary measures of teacher effectiveness, each with advantages and disadvantages. Measures of effectiveness based on peer observation circumvent some of the concerns that VAM raise, but present others. They are applicable to all teachers and could address a wide set of attributes in a teacher's practice. Also, they can be based on well-tested and widely used rubrics, so they are amenable to empirical analysis.

However, the need for evaluators raises issues of inter-rater reliability. Also, they evaluate teacher's practice, not student outcomes, and, they are limited slices of teacher behavior which may not represent typical behavior during the majority of time when they are not being observed (i.e. measure 'maximum' rather than 'typical' performance).

Student surveys, on the other hand, address both student outcomes and teacher practice, but from the viewpoint of the students. Development, testing and large scale application of student surveys as a basis for assessing teacher effectiveness is relatively new and far scarcer. Students have ongoing contact with teachers and have an ideal opportunity to observe their teaching. However, their involvement in the educational process means they are not objective observers. Student ratings may be contaminated by irrelevant experiences and factors beyond the scope of what policy-makers value.

Although a literature that focuses on multi-tiered teacher evaluations has started to emerge, it is sparse. The relationship between VAM and evaluations based on in-class observations has only been examined relatively recently (e.g. Kane et al, 2011; Rockoff & Speroni, 2010), while the use of multiple measures that also include feedback from student surveys is even more rare (e.g. Mihaly et al., 2013).

Teacher observations are generally correlated with VAM (e.g. Kane et al, 2011; Grossman et al., 2013). However, in multivariate regressions both predict future student success with little attenuating effects across coefficients, implying that they capture different aspects of teacher effectiveness (Rockoff & Speroni, 2010).

Notably, recent work based on the Gates Foundation's Measures of Effective Teaching (MET) Project (Kane & Staiger, 2012; Kane et al., 2013) explore how various indicators of teacher performance, including student surveys, in-class observation, and VAM might be used to build more complex measures of teacher effectiveness. Mihaly et al. (2013) found that there is a common component of effective teaching shared by all indicators, but there are also substantial differences across measures. The study finds that composite measures have higher stability over time than value added measures alone, but may do little to improve predictive power on future test scores.

Figure x Exit Patterns of All Teachers and Art, Language, Literature and Humanities Teachers

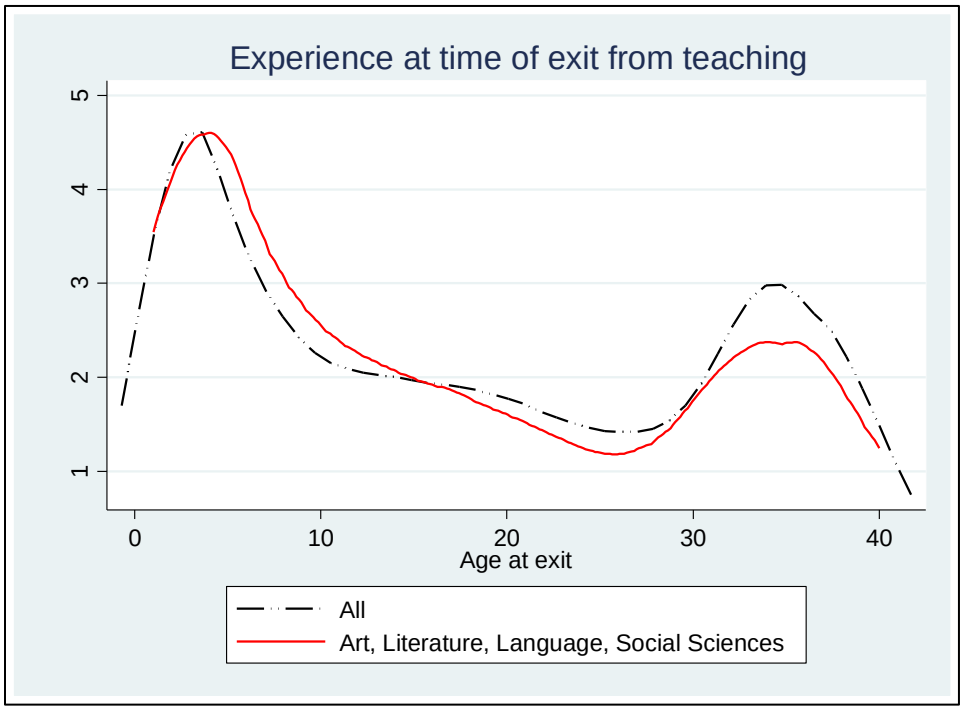


Figure x Exit Patterns of Elementary vs. STEM Teachers

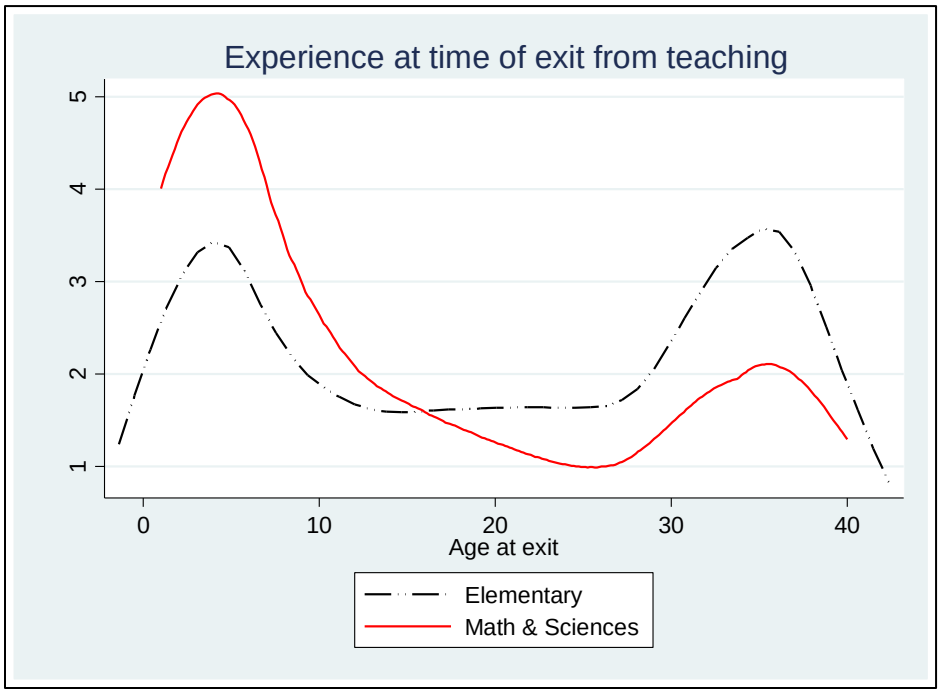


Figure 1.a Percent of Adults (25+) without a High School Degree

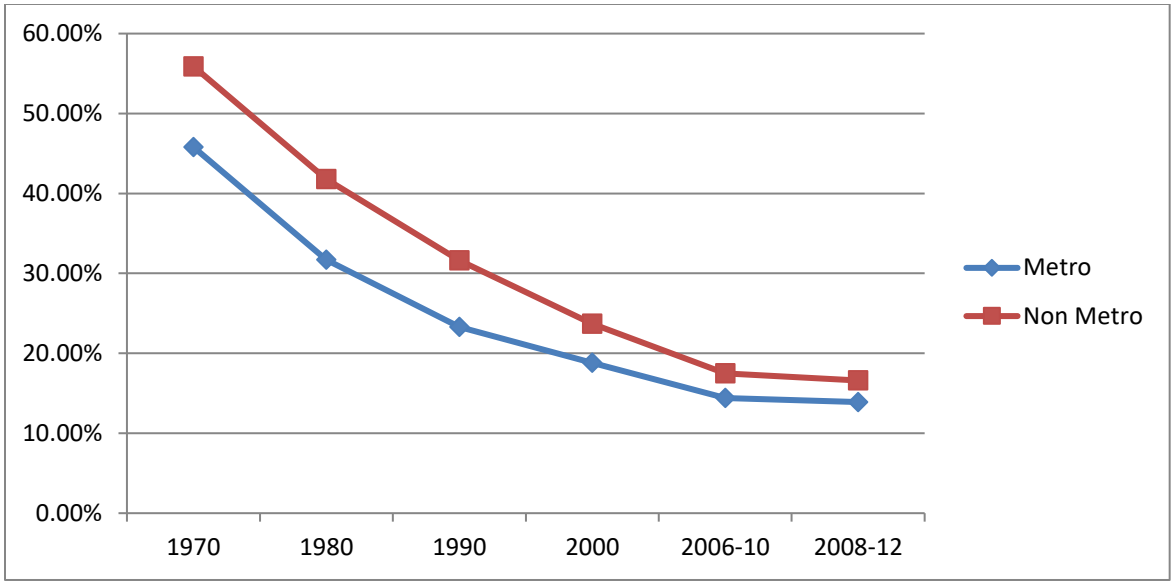


Figure 1.b Percent of Adults (25+) with at least a BS Degree

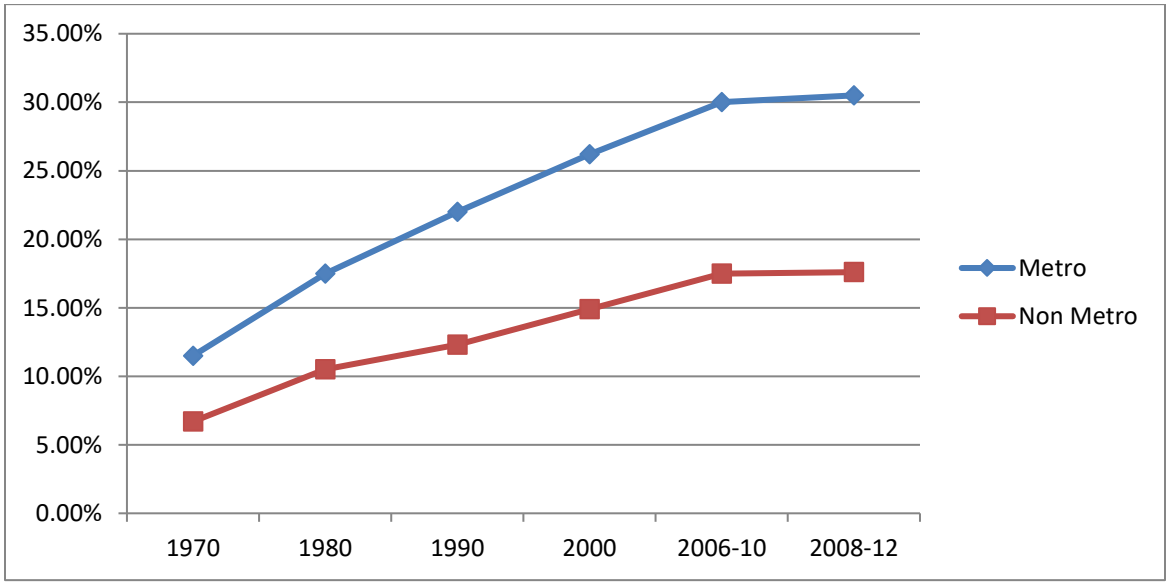
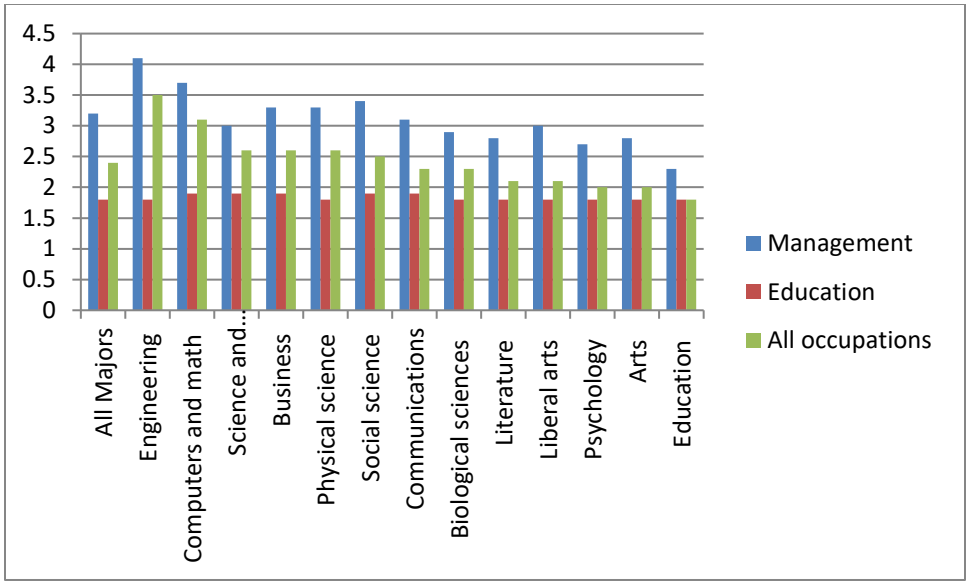


Figure 2 Lifetime Earnings By Major and Occupation



Source: US Census Bureau. American Community Survey (2011)

Sources: Decennial Census Summary Files (1970, 1980, 2000). American community survey 5-year aggregates (2006-10, 2008-12).

Andrews, Matthew, William Duncombe, and John Yinger. "Revisiting economies of size in American education: are we any closer to a consensus?." *Economics of Education Review* 21, no. 3 (2002): 245-262.

Cowen, Joshua M., J. S. Butler, Jacob Fowles, Megan E. Streams, and Eugenia F. Toma. "Teacher retention in Appalachian schools: Evidence from Kentucky." *Economics of Education Review* 31, no. 4 (2012): 431-441.

Jordan, Jeffrey L., Genti Kostandini, and Elton Mykerezi. "Rural and Urban High School Dropout Rates: Are They Different?." *Journal of Research in Rural Education* 27, no. 12 (2012): 1-21.

Mykerezi, Elton, Jeffrey L. Jordan, Genti Kostandini and Ilda Melo "On Rural-Urban Differences in Human Capital Formation: Finding the Bottlenecks" *Journal of Rural Social Sciences*, 29:1 (2014): 17–47.

Dolton, Peter, and Wilbert Van der Klaauw. "The turnover of teachers: A competing risks explanation." *Review of Economics and Statistics* 81, no. 3 (1999): 543-550.

Israel, Glenn D., Lionel J. Beaulieu, and Glen Hartless. "The Influence of Family and Community Social Capital on Educational Achievement*." *Rural sociology* 66, no. 1 (2001): 43-68.

Kusmin, Lorin D. “Rural America At A Glance”. No. EB-24. United States Department of Agriculture, Economic Research Service, (2013).

Appendix Table 1. Earnings by Major and Occupation (Nationally Representative Sample of College Graduates)

		All Majors	Engineering	Computers and Math	Science & Engineer Related	Business	Physical Science	Social Science	Biological Sciences	Lit.	Liberal Arts	Psych.	Arts	Educ.
1	All occupations	2.4	3.5	3.1	2.6	2.6	2.6	2.5	2.30	2.10	2.10	2.00	2.00	**1.80**
2	Architecture & engineering	3.4	3.6	3.3	2.9	2.9	2.9	2.6	2.80			3.50	2.80	**2.60**
3	Computer & math	3.2	3.7	3.4	3	3	3	3.2	3.00	2.90	2.90	2.80	2.70	**2.60**
4	**Management**	**3.2**	**4.1**	**3.7**	**3**	**3.3**	**3.3**	**3.4**	**2.90**	**2.80**	**3.00**	**2.70**	**2.80**	**2.30**
5	Business & finance	2.7	3.5	2.9	2.6	2.8	2.8	2.8	2.60	2.5	2.6	2.4	2.3	**2.1**
6	Healthcare	2.6			2.7	2.3	2.3	2.3	2.40	2.4	2.2	2.3	2.2	**2.1**
7	Sales and related	2.5	3.3	2.6	2.6	2.7	2.7	2.7	2.40	2.2	2.2	2.4	1.9	**1.9**
8	Science	2.5	2.9		2.5	2.4	2.7	2.4	2.40		2.80			
9	Arts and media	2.3	3	2.4	2.4	2.4		2.7	2.20	2.40	2.40		2.20	**1.90**
10	Production, etc.	1.9	2.4	2.2	1.7	1.9	2	1.7	2.00	1.40	1.60	1.70	1.70	**1.60**
11	**Education**	**1.8**	**1.8**	**1.9**	**1.9**	**1.9**	**1.8**	**1.9**	**1.80**	**1.80**	**1.80**	**1.80**	**1.80**	**1.80**
12	Community service & legal	1.8	3.2		2	1.9	1.8	1.8	2.20	2.00	1.70	1.70	1.80	**1.80**
13	Office support	1.7	2.1	1.8	1.8	1.8	1.8	1.8	1.80	1.60	1.60	1.60	1.60	**1.50**
14	Service	1.7	1.4	1.5	1.5	1.6	1.8	2	1.70	1.40	1.60	1.70	1.30	**1.30**

Source: American Community Survey, US Bureau of Census.