Journal of Research in Rural Education, Fall, 2003, Vol. 18, No. 2, 67-77

Evaluating Rural Progress in Mathematics Achievement: Threats to the Validity of "Adequate Yearly Progress"

Jaekyung Lee
SUNY at Buffalo

This article examines major threats to the validity of Adequate Yearly Progress (AYP) in the context of rural schools. Although rural students and their schools made significant academic progress in the past on national and state assessments, the current goal of AYP turns out to be highly unrealistic for them unless states set far lower achievement levels than those of the national assessment, NAEP. The current estimate of school AYP measures also turns out to be highly unreliable for small, rural schools, and using an uniform averaging procedure with multiple years of data would have only limited effects on stabilizing the measures. Finally, the current AYP formula that sets a uniform performance target for every school regardless of its initial performance status can unduly overidentify disadvantaged rural schools in need of improvement. Policy implications of the findings and possible ways to control those threats are discussed.

The No Child Left Behind Act (NCLB) requires standards-based accountability for school districts and schools receiving Title I funds. The intent of this law is to assure that all public schools accomplish academic excellence for all students. One major component of this accountability policy is to evaluate whether the districts and schools in each state are making "adequate yearly progress" (AYP) based on academic performance goals set by the state. This requirement applies to not only the entire body of students in each school but also to every subgroup of its students as broken down by their major demographic characteristics such as race/ethnicity and poverty.

What makes rural schools unique for evaluating their AYP? Research suggests that rural schools are in many ways different from urban schools (Howley, 1994). Almost 20% of the nation's rural schools have enrollments of fewer than 100 students (Stern, 1994). Small, rural schools tend to have a relatively lower percentage of minority students: Native Americans are overrepresented, but Asian and African American students are highly underrepresented (NCES, 1997). At the same time, the percentage of students in poverty tends to be higher in rural areas than in nonrural areas (NCES, 1997). Rural schools typically lack the facilities, physical plants, course offerings, and educational programs that typify larger, more resource-rich districts. In addition, rural schools face a major problem in recruiting and retaining qualified teachers and support personnel.

An earlier version of this paper was presented at the 2002 ACCLAIM Research Symposium on Rural Education. The author thanks the *JRRE* editor and the three anonymous reviewers for their constructive comments and suggestions.

Correspondence concerning this article should be addressed to Jaekyung Lee, SUNY at Buffalo, Graduate School of Education, 409 Baldy Hall, Buffalo, NY 14260. (jl224@acsu.buffalo.edu)

Previous studies have pointed out critical problems with present AYP-related efforts foreshadow technical challenges ahead (Hill, 1997; Lee & Coladarci, 2002; Linn & Haug, 2002; Thum, 2002). While most research and media tend to discuss the issues of AYP in general, little attention has been paid to the uniqueness of rural schools in this regard (however, see Coladarci, 2003). What are the implications of the NCLB's performance-based accountability policy mandate for rural schools and their students? What are the unique measurement challenges that small, rural schools would face in response to the policy mandate of meeting their AYP targets?

In light of these concerns, I examine how AYP, the linchpin of the NCLB's school accountability system, works for rural versus nonrural schools through analyses of national and state student assessment data.[1] A stated goal of NCLB is to provide accurate and meaningful information to the public on the quality of public schools and to identify schools in need of improvement. Therefore, the validity of this accountability system goes beyond simple validation of assessment tools used and depends upon how well the con-

[1] Schools are classified according to the Census Bureau definitions of metropolitan statistical areas (MSAs), population size, and density. Rural includes all places and areas with a population of less than 2,500. A Small Town is defined as places outside MSAs with a population of less than 25,000 but greater than or equal to 2,500. In this paper, schools in Central City, Urban Fringe, or Large Town are classified as "nonrural," and schools in Rural or Small Town as "rural." While rural areas in different states may have different social and cultural characteristics despite their common classifications, the Census Bureau definitions do not take into consideration the type of employment in that area and the degree of isolation. (See Khattri, Riley, & Kane, 1997 for different definitions of "rural.")

Table 1
Percentage of Eighth Graders Participating in School Lunch Programs and Their Average Mathematics Scores on the 1996 NAEP by Type of School Location

	Rural		Nonrural	
	Percentage	Average Score	Percentage	Average Score
Participated	32 (3.0)	261 (3.0)	37 (2.4)	247 (1.8)
Did not participate	68 (3.0)	282 (1.5)	63 (2.4)	277 (2.4)

Note. The NAEP mathematics scale ranges from 0 to 500. The standard errors of the statistics appear in parentheses. It can be said with about 95% confidence that for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample.

struction, use, and interpretation of AYP accomplishes this goal (see Marion, White, Carlson, Erpenbach, Rabinowitz, & Sheinker, 2002). Three major threats to the validity of AYP are identified and discussed in the context of rural schools.

First, is the AYP target feasible? If the AYP target dooms a majority of schools to fail, it will threaten the validity. Second, is the AYP measure reliable? If the AYP measure changes too much by chance, it will lose validity. Finally, is the AYP formula fair? If the AYP formula treats unequal schools equally, it will render the outcomes inequitable and lessen the validity. All of these questions are examined with a focus on rural schools and their students in the mathematics field. Using the National Assessment of Educational Progress (NAEP) results for the nation and NAEP and state assessment results for Maine, each of these three questions is addressed in the following sections.

Is AYP Feasible?

Since the passage of NCLB, much concern has been raised about the unrealistic AYP goal and timeline (i.e., 100% of students become proficient within 12 years) and its possible consequences for schools that repeatedly fail to meet AYP. It was estimated that up to 80% of schools in some states could be targeted as needing improvement or corrective action in the first few years (Olson, 2002, April 18). Further, the "safe harbor" provision does not help schools very much (see Lee, in press). However, it remains to be shown whether and how the feasibility of AYP varies by school location (urban, suburban, rural).

While NAEP provides a nationally representative portrait of rural vs. nonrural student achievement, it is difficult to track academic progress consistently over time due to changes in NAEP's location classification system.[2] Applying the same location classification, I was able to examine rural vs. nonrural progress in NAEP eighth grade math achievement for the nation and states between 1992 and 1996 (Lee, 2001; Lee & McIntire, 2001). This research showed that the most significant improvement occurred in rural schools nationwide between 1992 and 1996. In 1996, rural students outperformed nonrural students on the NAEP eighth grade mathematics assessment by approximately one fourth of a standard deviation.

The percentage of rural students at or above the NAEP proficient level in eighth grade mathematics increased from 17% to 25% between 1992 and 1996 (see Figure 1). If we use the 1992 and 1996 measures as the basis of projection and assume that the nation will make the same amount of gain (i.e., 8%) every 4 years after 1996, we can project that rural students at or above proficient will increase to 53% by 2014. This figure stops well short of the goal of 100% proficient. Even with the same amount of continuous gain since this time, it might take another 24 years to reach the 100% target. These projections are likely to be gloomier for nonrural students who were not able to make significant gains between 1992 and 1996. If we assume the nation adopts a lower achievement level as its target ("basic" instead of "proficient"), the deficit would be smaller, but many nonrural students might remain below the goal in 2014.

As NCLB requires subgroup reporting of academic progress, closing the achievement gap among subgroups

[2] Traditionally, NAEP has reported math proficiency by a community-type variable (advantaged urban, disadvantaged urban, and extreme rural) that combined community size with a school-level socioeconomic indicator. Discontinuing this classification due to the problematic nature of the variable, NAEP began reporting results by Census-based location type since 1996. In the 2000 NAEP, the same type of location variable was used, but it was not comparable to past data. This was due to the fact that a new method was used to identify the location type assigned to each school in the Common Core of Data (CCD); schools were not classified in exactly the same way in 2000 as in previous years in terms of location type (Braswell et al., 2001).

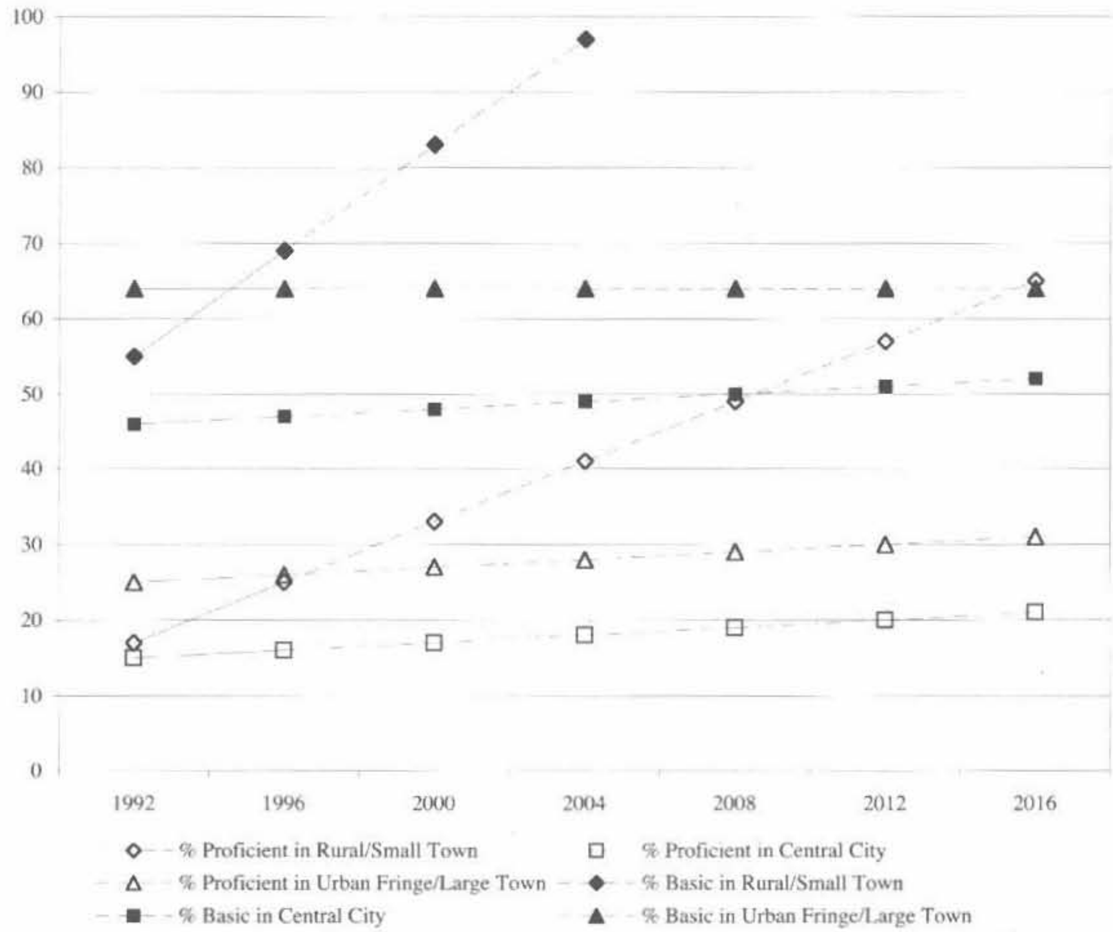

Figure 1. Trajectories of percent students at or above basic and proficient achievement levels based on the 1992 and 1996 NAEP eighth grade math assessment results by type of location (actual measures are shown in solid lines and projected estimates are shown in broken lines)

has become a major challenge for both rural and nonrural schools. Table 3 shows the 1996 NAEP math achievement of eighth grade students as classified according to whether they participated in school lunch programs (as an indicator of poverty) in rural and nonrural schools nationwide. There was no significant difference between rural and nonrural schools in terms of the percentage of free or reduced-price lunch students. In both rural and nonrural schools, students who participated in school lunch programs scored lower than their nonparticipating classmates. However, the achievement gap between participants and nonparticipants is much more significant in nonrural schools than in rural schools. More importantly, free or reduced-priced lunch students in rural schools performed significantly better than their counterparts in nonrural schools. This indicates that rural schools might be in a better position than nonrural schools for meeting AYP targets.

Despite these national trends, substantial variations exist among states in rural students' mathematics achievement and their achievement gains (Lee, 2001, 2002; Lee & McIntire, 2001). First, some rural states performed at the top (e.g., Iowa, Maine), while others performed below the national average (e.g., Arkansas, Mississippi). Second, there are interstate variations in rural students' mathematics achievement gain over the 1992-1996 period (see Figure 2). Among the 35 states participating in the 1992 and 1996 NAEP eighth grade mathematics assessments, rural students made statistically significant progress in 12 states (Florida, Kentucky, Maryland, Michigan, Maine, New Mexico, North Carolina, Tennessee, Texas, Utah, Wisconsin, and West Virginia). For those states, the size of their gain is small to moderate, ranging from 4 to 12 points (approximately .1 to .4 standard deviations).

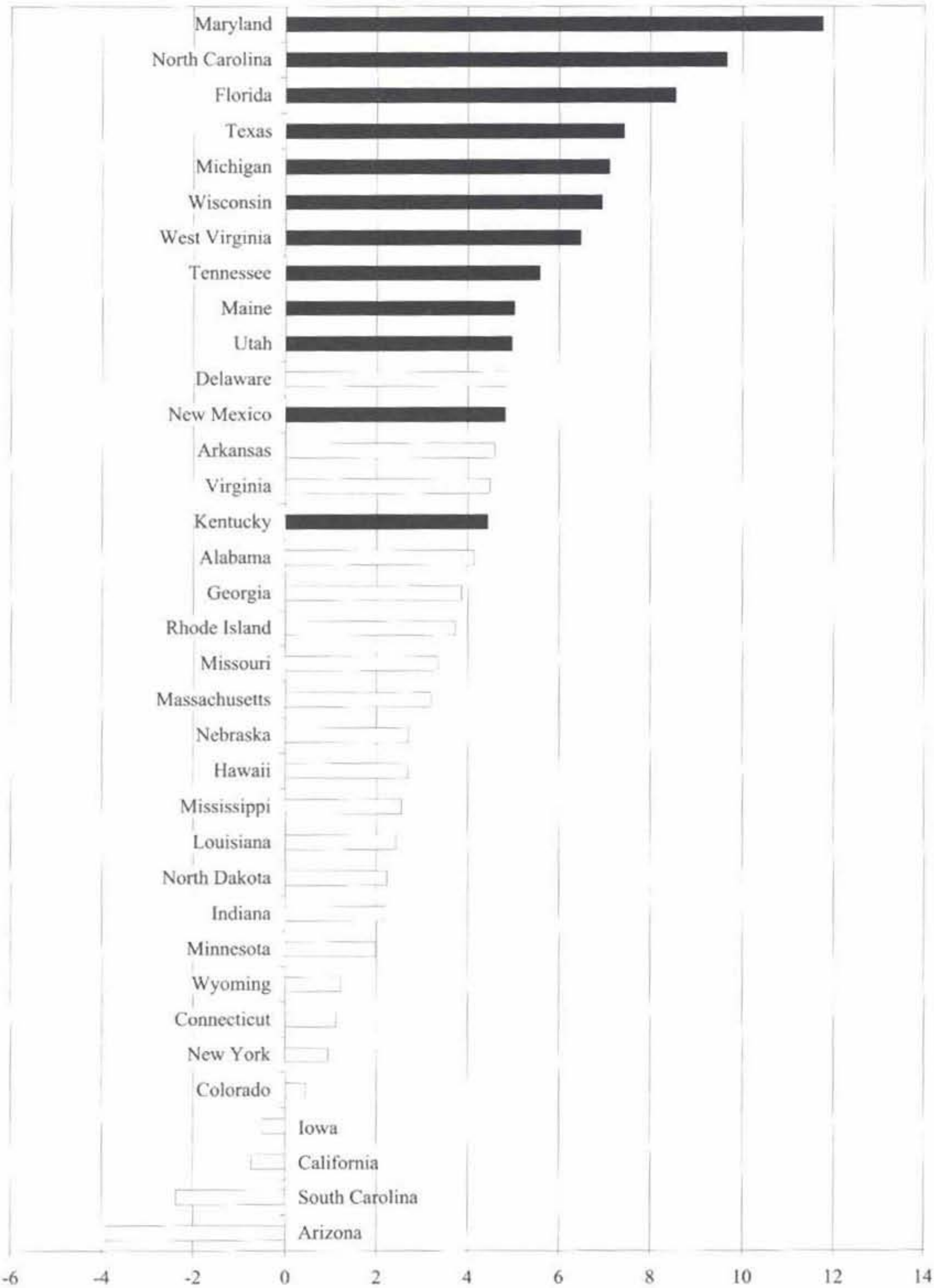

Figure 2. State average eighth grade NAEP math achievement gains from 1992 to 1996 in rural/small towns ($N = 35$ states; states with statistically significant gains are shown in black bars)

Table 2
Maine Eighth Grade Average Mathematics Scores by Type of School Location on MEA and NAEP, 1992 and 1996

Assessment	Locale	1992	1996	Raw Gain	Standardized Gain
MEA	Nonrural	315	346	31*	0.23
	Rural	296	343	47*	0.36
NAEP	Nonrural	282	288	6	0.19
	Rural	278	283	5*	0.16

Note. Asterisk indicates that the gain is statistically significant at the .05 level.

The percentage of proficient students on NAEP also shows uneven progress among states. For example, Maine—one of the highest performing and the most improving states, with the majority of its students living in rural areas—shows a modest increase in the percentage of proficient rural students: from 25% to 30% between 1992 and 1996. Assuming a 5% gain for every 4-year period, Maine will have 47.5% of its rural students meeting the proficient level by 2014. These data suggest that the current expectation regarding AYP can be highly unrealistic—"delusional," in the words of Coladarci (2003)—unless states set far lower achievement levels than those of NAEP. When we lower our achievement target to the NAEP "basic" level, the goal might become attainable. Maine rural students at or above basic in 2014 will be 98.5. As the current law insists on using "proficient" as its target achievement level, the appropriateness of current NAEP achievement levels as a benchmark for states needs to be evaluated. The previous evaluations raised serious concerns about the reliability and validity of the NAEP achievement levels and the possibility of misleading interpretations (see Shepard, Glaser, Linn, & Bohrnstedt, 1993; U.S. General Accounting Office, 1993).

Is AYP Reliable?

We should use multiple measures in a context of high stakes for students and/or their school systems (see American Educational Research Association, American Psychological Association, & National Council on Educational Measurement, 1999). This is also true for evaluation of school AYP. Using more than a single measure may allow us to get a more comprehensive picture of student achievement and enhance the reliability and validity of evaluation.

NCLB requires each state to participate biennially in NAEP assessments of fourth and eighth grade reading and mathematics. Similarly, NCLB requires each LEA to participate, if selected, in the state NAEP. The law does not require that NAEP results confirm progress on state tests, but its mandate that all 50 states now take part in the NAEP makes such comparisons more likely (Olson, 2002, March 13).

Are achievement gains as measured by a state's own assessment valid? While NAEP may be used as a tool for the U.S. Department of Education to cross-check and validate state-level or district-level academic progress, previous comparisons of NAEP and state assessment results showed significant discrepancies in the size of statewide achievement gains (Lee & McIntire, 2002). However, it remains to be shown whether this problem applies to both rural and nonrural schools and how much rural progress reported by states differs from rural progress based on the NAEP.

Table 2 compares Maine student performance improvement levels based on the NAEP and Maine Educational Assessment (MEA) eighth grade math assessment results by location type.[3] As shown in Table 2, we find that both rural and nonrural students in Maine made academic improvement between 1992 and 1996 as measured by the MEA and NAEP. However, the size of state gains in math tend to be somewhat greater than on the NAEP: approximately 2 times larger for rural students, and 1.25 times larger for nonrural students. This does not tell us which assessment is more accurate and valid, but it shows that choosing a particular assessment can make a difference. One plausible reason for this difference is that the NAEP sample size of rural students is relatively smaller and thus its estimate of rural student achievement entails larger standard errors. Further research is needed to find out whether the discrepancy between national and state assessment results on academic progress is larger for rural students than for nonrural students in other states as well.

School-level achievement gains tend to be much less reliable than state-level achievement gains. The current measure of school AYP is based on comparison of successive student groups' performance at the same grade level and tends to be highly unreliable (Kane & Staiger, 2002; Lee & Coladarci, 2002; Linn & Huag, 2002). This lack of reliability is more serious for small, rural schools when the law

[3] Because NAEP and MEA scores employ different scales, I established a common metric in standard deviation units. Specifically, I used student standard deviations from the MEA 1996 mathematics assessment results to compute MEA standardized gain, while I used Maine's standard deviations from the 1996 NAEP state assessment results to compute NAEP standardized gain.

Table 3

Average Number of Years in which Maine Schools Would Have Met AYP Target in Eighth Grade Mathematics During 1990-1998 Period Under "Uniform" vs. "Individualized" AYP-Setting Approaches by School Performance Level and Location Type

Performance	Locale	Uniform	Individualized
Low	Nonrural	4.0	6.5
	Rural	4.4	6.2
High	Nonrural	8.3	5.3
	Rural	7.4	4.8

Note. Low-performing schools are the schools whose eighth grade math score was initially below the AYP target in baseline year 1990. High-performing schools are the schools whose eighth grade math score was initially at or above the AYP target in baseline year 1990.

requires reporting the progress of every major demographic subgroup in each school (e.g., see Coladarci, 2003). Wrong identification of schools with needs for academic improvement and the following misallocation of resources/aids may result from the use of unreliable AYP measures.

Wheat (2000) argued that states should switch from this successive cohort model assessing different cohort groups' achievement to a longitudinal growth model assessing the same cohort's achievement gains. For instance, schools may be evaluated by comparing this year's 3rd grade passing rates (i.e., percentage of students at or above proficient level) with next year's fourth grade passing rates achieved by the same cohort of students. Although Hill (2001) has shown this longitudinal evaluation method is more reliable, it demands that states track individual students' academic performance over time with comparably rigorous performance standards across different grades. Mobility rates of students could also limit a state's choice of this approach, because longitudinal approaches would exclude large numbers of students in schools with high mobility (Marion et al., 2002). While rural districts tend to have a high student mobility rate— approximately the same as the national average of 15, state and local education agencies are increasingly being held accountable for the education of highly mobile students (Paik & Phillips, 2002).

A state department of education web site usually provides publicly available information on school achievement that can be used for AYP purposes. State assessment data can be used in combination with the Common Core of Data (CCD) that contains Census-based information on the type of school location and is available from the National Center for Education Statistics. This combination of data allows one to compare academic progress in rural vs. nonrural schools in particular states. The following analysis of Maine schools' eighth grade math achievement gains using the MEA data collected during the 1990-1998 period and 2000 CCD data shows that the estimation of academic progress in small, rural schools based on the successive cohort comparison method is highly unreliable. Thus, the current AYP formula and the allocation of resources to schools based on such unreliable AYP measures can be misleading.

Figure 3 illustrates two randomly selected schools in Maine (one from a rural area and another from a nonrural area). This particular rural school shows enormous volatility in its average math achievement score throughout the 1990-1998 period, which makes it hard to detect its overall performance trend. In contrast, this nonrural school shows a high level of stability with its generally upward performance trend. This difference in performance trend reflects the schools' differences in size and enrollment trends. The rural school tended to have fewer than 40 students tested, and this number changed substantially over time: 12 in 1995, 37 in 1996, 22 in 1997, and 37 in 1998. The nonrural school had more than 200 students tested and the number was relatively stable: 245 in 1995, 241 in 1996, 254 in 1997, and 255 in 1998.

Many small elementary schools that test only in one grade currently (e.g., fourth) could double, triple or quadruple their effective sample size when third, fifth, and possibly sixth grade (depending on the school configuration) results are included in AYP analyses. The same can be said of middle schools that test only eighth grade now but may increase the sample size as many as four times once state assessments of grades 5-7 are put in place. If a school increases its effective testing sample size four times, it may reduce the school's standard error of AYP measure up to half.

Under the NCLB AYP provisions, schools have the option of "using a uniform averaging procedure, which is designed to mitigate the fact that student performance can vary widely from year to year due to factors beyond a school's control" ("Raising the Bar," 2002). Under this provision, schools can average test scores from the current school year with test scores from the preceding two years. In order to see how well this provision can reduce variations in school performance trends, I examined the variability of Maine schools' MEA eighth grade average mathematics scores over 1990-1998 period with and without an application of this rolling average procedure. Figure 4 shows the

Figure 3. Two sample Maine schools' eighth grade MEA math performance trajectories

distributions of rural vs. nonrural schools' score variability as measured by the standard deviations of their 9-year eighth grade average math scores. For an average nonrural school, the variability of mathematics scores drops from 30 to 20 in standard deviation units by using this rolling average. For an average rural school, the same measure of variability decreases from 40 to 30. This comparison tells us that rural schools still remain relatively more unstable than their nonrural counterparts, as the use of rolling average procedure tends to reduce the variability by the same degree for both nonrural and rural schools.

It is particularly challenging to measure achievement gains made by even smaller subgroups of students. Even in relatively large rural schools, the number of racial and ethnic minority students is likely to be much smaller than that in urban schools. There are provisions to exempt states from the requirement to report disaggregated data when the number of students in a category is insufficient to yield statistically reliable information or the results would reveal personally identifiable information about an individual student. However, different states may develop different criteria about the minimum number of students to address this issue, since the U.S. Department of Education does not provide specific guidelines (MacQuarrie, 2002). While a higher minimum number yields fewer schools failing AYP, it has been shown that more valid and reliable AYP decisions can be made by using confidence intervals (Marion et al., 2002; also see Coladarci, 2003). It remains to be seen how the states' choices will differentially affect rural schools.

Is AYP Fair?

The current AYP policy directs its support and accountability (i.e., aids and sanctions) to relatively low-performing schools that receive Title I funds, but it does not take into account disparities among different schools in their capacity to reach the goal. NCLB gives rural schools greater flexibility in using federal funds for improving student learning. The Small, Rural School Achievement Program is designed for small or rural districts that frequently lack personnel and resources (New England Center for Educational Policy and Leadership, 2002). There are, however, many poor small rural schools for which existence is potentially threatened by corrective actions and sanctions when they fail to meet state-imposed AYP targets.

The final U.S. Department of Education regulations do not permit calculating the starting points separately for any subgroups. By setting uniform AYP targets for every school and every student subgroup, the current formula does not consider the influence of schools' initial achievement status

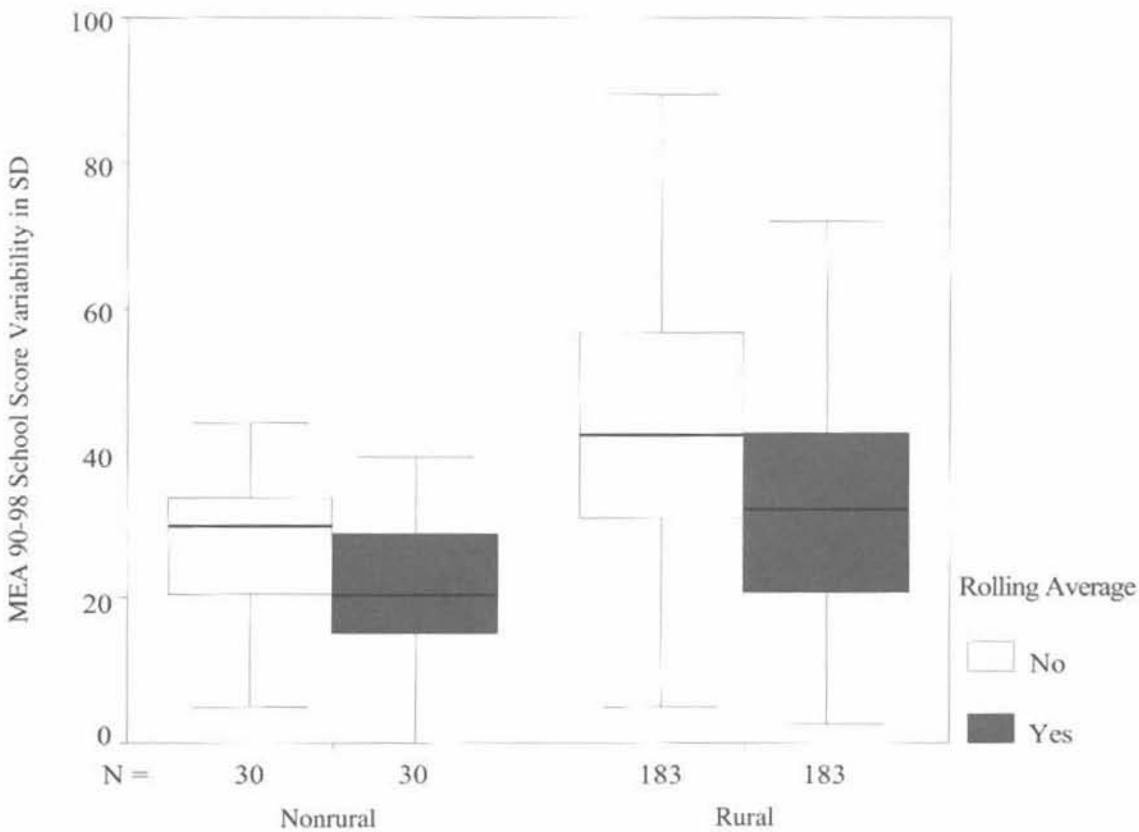

Figure 4. Rural vs. nonrural Maine schools' variability of 1990-98 MEA eighth grade mathematics scores in standard deviation units with and without rolling average procedure (N = 30 nonrural schools and 183 rural schools)

on their chance to meet the target consecutively. Higher-performing schools that are above the AYP target at the beginning will be able to meet the target more easily than lower-performing schools that are initially below the target. Even among those lower-performing schools, schools that are closer to the target initially are in a better position to meet the AYP target continuously.

In contrast, the individualized AYP-setting approach that I would recommend would set different targets for schools predicated on the school's baseline status. In other words, every school would have its own AYP target, although the ultimate goal is the same for all schools. Schools that initially performed at a lower level would have to make relatively large gains, while initially higher-performing schools would be allowed smaller gains. If there is significant achievement gap between rural and nonrural schools in a state as attributable to disparities in their educational opportunities and resources, then the state may consider setting separate AYP target norms for rural vs. nonrural schools.

Figure 5 illustrates uniform vs. individualized approaches to setting AYP targets over a 12-year time period. Suppose School A was initially performing above the statewide uniform AYP target (thick solid line) while School B was initially performing below the same target. Although both schools may ultimately reach the goal of 100% proficient in 12 years, they are likely to take quite different paths to the goal. If each of the two schools follows its projected linear performance trajectories (thin lines), School A would meet the target throughout the 12-year period, whereas School B would never meet the target except year 12. This result does not make sense because school B made much greater progress than School A throughout the period. In order to avoid becoming a failing school, School B would have to increase its performance substantially during the first year to reach its target. If AYP targets are individualized following expected growth trajectories for each school, both School A and School B would be able to meet AYP targets more equally.

To test how different AYP-setting approaches work, I analyzed actual data from Maine schools' 1990-1998 MEA eighth grade math and compared the results from an individualized AYP-setting method with that from a uniform AYP-setting method (i.e., current AYP formula). For testing a uniform AYP approach, the current AYP formulas

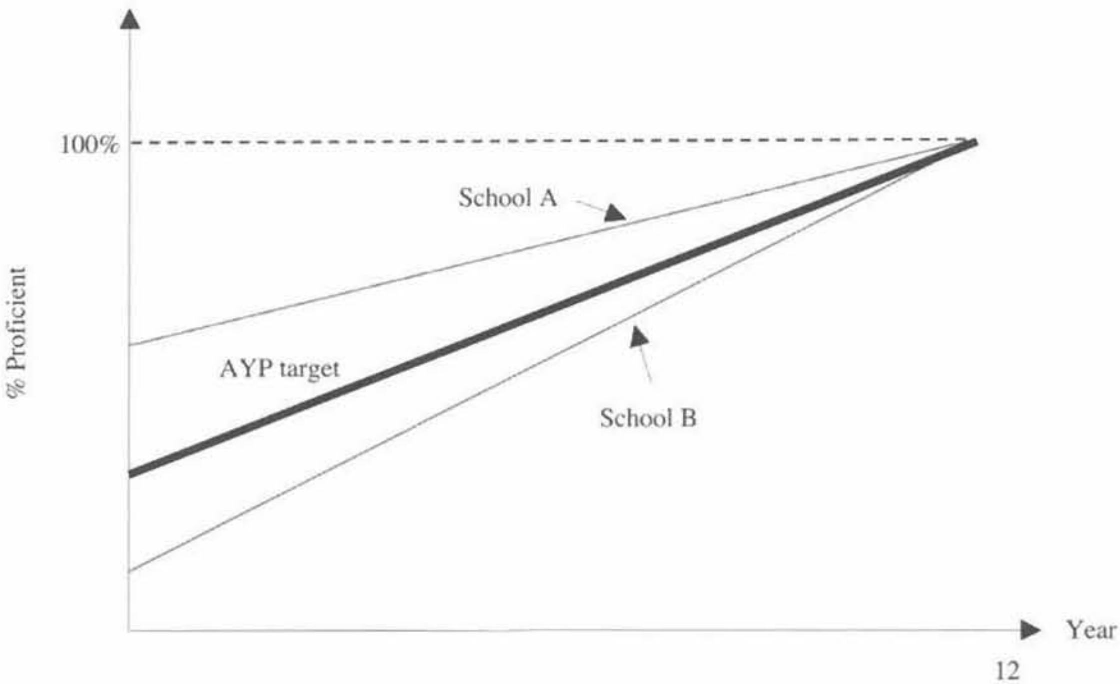

Figure 5. Uniform AYP targets for all schools (thick line) vs. individualized AYP targets for two sample schools (thin lines), both aimed at the goal of 100% proficiency within 12-year timeline

were used to determine baseline and annual AYP targets in Maine: Average score at the state's 20th percentile school was treated as baseline (score of 250) in 1990, and above that baseline equal increments (10 points) were made every year so that the AYP target became 370 in 2002 (12 years from 1990). This hypothetical AYP target was a reasonably high standard, as 370 is about 2 standard deviations above the starting year's statewide average and Maine schools made about an 8-point gain on average per year during the 1990-1998 period. For testing an individualized AYP approach, each school's baseline target was set at its 1990 score and increments were made individually according to its projected performance trajectories.

Table 3 shows how many times rural and nonrural schools in Maine would have met the AYP target throughout the 1990-1998 period under this uniform AYP formula scenario. Low-achieving schools that were initially below the 20th percentile rank would have met the AYP target in only 4 years out of 9 years total, while schools that were initially at or above the 20th percentile rank would have met the target about twice as many times. (Both rural and nonrural schools followed similar patterns.) Table 3 also shows the results of applying the individualized AYP formula to the same data. Now schools that were initially below the 20th percentile would have met the AYP target five times, while schools that were initially above the 20th percentile met the target six times over the 9-year period. Therefore, this alternative AYP-setting approach tends to produce significantly better AYP results for low-achieving schools and make the chance of their meeting AYP target as high as their high-achieving counterparts.

Conclusion

Because my analyses included only eighth grade mathematics achievement data collected during the 1990s, these results may not be directly applicable to other grades, subject areas, and time periods. As NCLB requires testing all students in each grade from 3 through 8 and once in grades 10 to 12 in reading and mathematics by 2006, the situation would become different than what states faced in the past. Indeed, projection of future trends based on past performance results might underestimate schools' potential progress expected under this new legislation. The results may have been different if schools had faced the stronger incentives embodied in current AYP rules. Moreover, combining data from multiple grades could produce more reliable estimates of school performance measures than relying on data from a single grade. With these caveats in mind, my results suggest that we become more aware of potential problems and challenges in evaluating academic progress made by rural schools and their students.

The most imminent challenge of measuring rural schools' and their students' achievement gains comes from the NCLB AYP mandate. Do rural schools and their students make adequate yearly progress? Do rural students perform as well as their nonrural counterparts? Are the achievement gaps among different socioeconomic groups of students in rural schools closing? Rural educational policymakers and practitioners might ask these kinds of questions that inevitably arise from the new legislation that requires regular evaluation and reporting of academic progress in core subjects, including mathematics.

Given that many rural students are poor and attend schools whose instructional resources and course offerings are limited, the level of their academic performance relative to their nonrural counterparts is encouraging. Indeed, rural schools, having achieved so much with relatively few resources, can provide a model of strength worth studying and emulating. Nevertheless, many disadvantaged rural schools across the nation are highly unlikely to reach the current AYP goal unless we lower the target achievement level or extend the timeline to reach the level. While rural schools appear to have a better prospect of progress than their nonrural counterparts based on the past NAEP and state assessment results, they, too, get very gloomy projections far behind their expected target AYP level. It remains to be seen whether they can even maintain the past progress under the current climate of school budget crunch.

Underlying measurement problems, including the lack of consistency and accuracy, also threaten the validity of current AYP measures. While these technical issues apply to both rural and nonrural schools, there are some unique aspects of the problems for rural schools. The findings of this study reveal differences between rural and nonrural schools. First, both rural and nonrural students showed greater achievement gains on state assessment than on the NAEP, but the difference tended to be larger for rural students. As the NAEP may provide a less accurate estimate of student achievement in rural areas due to its smaller sample size, we need extra cautions in interpreting the NAEP results for rural students. Second, the AYP measures were unstable for both rural and nonrural schools over the past years, but they turned out to be more unreliable for small, rural schools. As state assessment results for rural schools are highly vulnerable to change over time, we need special care in evaluating the AYP results of small, rural schools.

While those measurement issues are generally related to school size rather than location, there are also other unique aspects of being rural that may influence AYP results (e.g., racial homogeneity of students, geographic isolation, lack of instructional resources and advanced course offerings). However, all rural schools do not have the same academic needs. There are significant variations among states in the achievement level of rural students as well as within states. Since rural schools in isolated and poor communities have a greater risk of failing to meet state-imposed uniform AYP targets, it is not fair to apply the same criteria of evaluation to all rural schools across the nation and states in the same ways. While the individualized AYP-setting approach with different starting points for different schools tends to reduce the identification gap, it implies changes in the current law and raises controversies about the correct way to identify the most academically needy schools. In order to address this equity issue within the parameters of the current law and regulations, states still need to fully assess the unique needs of small, disadvantaged rural schools and provide them with special aids to an end that is more equitable.

References

American Educational Research Association, American Psychological Association, & National Council on Educational Measurement (1999). *Standards for educational and psychological testing.* Washington, DC: AERA.

Braswell, J. S., Lutkus., A. D., Grigg, W. S., Santapau, S. L., Tay-Lim, B., & Johnson, M. (2001). *The nation's report card: Mathematics 2000* (NCES Report No. 2001–517). Washington, DC: National Center for Education Statistics.

Coladarci. T. (2003). *Gallup goes to school: The importance of confidence intervals for evaluating "adequate yearly progress" in small schools.* Washington, DC: Rural School and Community Trust. (http://ruraledu.org/docs/nclb/coladarci.pdf)

Hill, R. (1997, June). *Calculating and reducing errors associated with the evaluation of adequate yearly progress.* Paper presented at the Annual Assessment Conference of the Council of Chief State School Officers, Colorado Springs, CO. (ERIC Document Reproduction Service No. ED414307)

Hill, R. (2001, October). *Issues related to the reliability of school accountability scores.* Paper presented at the Reidy Interactive Lecture Series, Nashua, NH. Retrieved from www.nciea.org

Howley, C. (1994). *The academic effectiveness of small-scale schooling: An update.* Charleston, WV: ERIC Clearinghouse on Rural Education and Small Schools.

Khattri, N., Riley, K. W., & Kane, M. B. (1997). Students at risk in poor, rural areas: A review of the research. *Journal of Research in Rural Education, 13,* 79-100.

Kane, T. J., & Staiger, D. O. (2002). Volatility in school test scores. In D. Ravitch (Ed.), *Brookings papers on education policy 2002* (pp. 235-284). Washington, DC: Brookings Institution.

Lee, J. (in press). How feasible is adequate yearly progress (AYP)? Simulation analyses of Maine and Kentucky School AYP Data. *Educational Policy Analysis Archives.*

Lee, J. (2002). Understanding differences between rural and nonrural schools impacting student achievement: Implications for title I programs and assessments. In

K. K. Wong & M. C. Wang (Eds.), *Accountability and efficiency in title I schoolwide reform* (pp. 291-310). Greenwich, CT: Information Age Publishing.

Lee, J. (2001). *Interstate variations in rural student achievement and schooling conditions* (ERIC Digest). Charleston, WV: ERIC Clearinghouse on Rural Education and Small Schools. Retrieved from http://www.ed.gov/databases/ERIC_Digests/ed459037.html

Lee, J., & Coladarci, T. (2002). *Using multiple measures to evaluate the performance of students and schools: Learning from the cases of Kentucky and Maine.* Orono, ME: University of Maine. (http://www.ume.maine.edu/naep/SSI)

Lee, J., & McIntire, W. (2002). *Using national and state assessments to evaluate the performance of state education systems: Learning from the cases of Kentucky and Maine.* Orono, ME: University of Maine. (http://www.ume.maine.edu/naep/SSI)

Lee, J., & McIntire, W. (2001). Interstate variation in the mathematics achievement of rural and nonrural students. *Journal of Research in Rural Education, 16*(3), 168-181.

Linn, R. L., & Haug, C. (2002). Stability of school-building accountability scores and gains. *Educational Evaluation and Policy Analysis, 24*(1), 29-36.

MacQuarrie, D. (2002, September). The No Child Left Behind Act: Regulatory guidance. *NCME Newsletter 10*(3), 3-4.

Marion, S. F., White, C., Carlson, D., Erpenbach, W. J., Rabinowitz, S., & Sheinker, J. (2002). *Making valid and reliable decisions in the determination of adequate yearly progress.* Washington, DC: Council of Chief State School Officers.

National Center for Education Statistics. (1997). *Characteristics of small and rural school districts.* Washington, DC: U.S. Department of Education.

New England Center for Educational Policy and Leadership. (2002). *Implementing the No Child Left Behind Act of 2001: A tool kit for New England state policy makers.* Storrs, CT: Author.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002). (http://www.ed.gov/legislation/ESEA02/)

Olson, L. (2002, March 13). Want to confirm state test scores? It's complex, but NAEP can do it. *Education Week.*

Olson, L. (2002, April 18). 'Inadequate' yearly gains are predicted. *Education Week.*

Paik, S. J., & Phillips, R. (2002). *Student mobility in rural communities: What are the implications for student achievement?* (NCREL Report). Retrieved from http://www.ncrel.org/policy/pubs/html/rmobile/index.html

Raising the bar: The complexities of "adequate yearly progress". (2002). *Education Assessment Insider 5, 1*(1).

Shepard, L. A., Glaser, R., Linn, R. L., & Bohrnstedt, G. (1993). *Setting performance standards for student achievement: An evaluation of the 1992 achievement levels.* Stanford, CA: National Academy of Education.

Stern, J. (Ed.). (1994). *The condition of education in rural schools.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Thum, Y. (2002). *Design of school performance and school productivity indicators: Measuring student and school progress with the California API.* Unpublished manuscript.

U.S. General Accounting Office (1993). *Educational achievement standards: NAGB's approach yields misleading interpretations* (GAO/PEMD-93-12). Washington, DC: Author.

Wheat, D. (2000). *Value-added accountability: A systems solution to the school accreditation problem* (2nd ed.). Springfield, CA: Thomas Jefferson Institute for Public Policy. (http://www.thomasjeffersoninst.org/pdfs/VAA_2nd_edition.pdf)