DOCUMENT RESUME

ED 475 060 RC 023 789

AUTHOR	Lee, Jaekyung
TITLE	Evaluating Rural Progress in Mathematics Achievement: Is "Adequate Yearly Progress" (AYP) Feasible, Valid, Reliable, and Fair? Working Paper.
INSTITUTION	Ohio Univ., Athens. Appalachian Collaborative Center for Learning, Assessment, and Instruction in Mathematics.
SPONS AGENCY	National Science Foundation, Arlington, VA.
REPORT NO	WP-11
PUB DATE	2003-03-00
NOTE	36p.; Paper presented at the ACCLAIM Research Symposium (McArthur, OH, November 3-6, 2002).
CONTRACT	NSF-0119679
AVAILABLE FROM	For full text: http://kant.citl.ohiou.edu/ACCLAIM/ rc/rc_sub/pub/3_wp/Lee11.pdf.
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	Academic Standards; *Accountability; Feasibility Studies; Grade 8; High Stakes Tests; *Mathematics Achievement; Mathematics Education; Middle Schools; Policy Analysis; *Reliability; *Rural Schools; School Size; Small Schools; Testing Problems; *Validity
IDENTIFIERS	Fairness; Maine; *No Child Left Behind Act 2001

ABSTRACT

The No Child Left Behind Act requires standards-based accountability for school districts and schools receiving Title I funds. A major component of this policy is to report whether districts and schools are making "adequate yearly progress" (AYP) based on their performance goals. This paper raises questions for rural schools using the National Assessment of Educational Progress (NAEP) mathematics scores from 35 states and state student assessment results from Maine between 1992 and 1996. Assuming the nation's rural students will make the same amount of gain every 4 years as they did between 1992 and 1996, the number of rural students at or above proficient in mathematics will rise to 53 percent by 2014, indicating that the AYP goal is not feasible. Overall statewide academic improvement in Maine was approximately 2 times larger using the state assessment than with the NAEP assessment. That the assessment used can make such a large difference raises questions of validity. Because smaller sample sizes inherently produce unreliable scores, the successive cohort comparison is highly unreliable as a measure of academic progress in small rural schools. By setting a uniform AYP target for every school, the current formula does not consider the influence of schools' initial performance status on their chance to meet the target, which brings the fairness of the AYP into question. Recommendations include lowering the target achievement level or extending the timeline to reach the level for disadvantaged schools, allowing the use of multiple measures to demonstrate school progress, using rolling averages to stabilize performance variations, and allowing individualized AYP targets according to baseline performance levels. (Contains 19 references.) (TD)

Appalachian Collaborative Center for Learning, Assessment and Instruction in Mathematics

Working Paper Series

Evaluating Rural Progress in Mathematics Achievement: Is "Adequate Yearly Progress" (AYP) Feasible, Valid, Reliable, and Fair?

Working Paper No. 11

Jaekyung Lee
SUNY Buffalo
March 2003

ACCLAIM's mission is the cultivation of *indigenous leadership capacity* for the improvement of school mathematics in rural places. The project aims to (1) understand the rural context as it pertains to learning and teaching mathematics, (2) articulate in scholarly works, including empirical research, the meaning and utility of that learning and teaching among, for, and by rural people, and (3) improve the professional development of mathematics teachers and leaders in and for rural communities.

OHIO
UNIVERSITY

ACCLAIM Research Initiative
All rights reserved

Address: 210A McCracken Hall
Ohio University
Athens, OH 45701-2979

Office: 740-593-9869
Fax: 740-593-0477

E-mail: howleyc@ohio.edu
Web: http://acclaim.coe.ohiou.edu/

Funded by the National Science Foundation as a Center for Learning and Teaching, ACCLAIM is a partnership of the University of Tennessee (Knoxville), University of Kentucky (Lexington), Kentucky Science and Technology Corporation (Lexington). Marshall University (Huntington, WV), University of Louisville, and Ohio University (Athens, OH).

ACCLAIM

This material is based upon the work supported by the National Science Foundation Under Grant No. 0119679. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Evaluating Rural Progress in Mathematics Achievement:

Is "Adequate Yearly Progress" (AYP) Feasible, Valid, Reliable, and Fair?

Jaekyung Lee

SUNY Buffalo

Paper presented at the
ACCLAIM Research Symposium
Ravenwood Castle, McArthur, Ohio

November 3-6, 2002

The reauthorized Elementary and Secondary School Act (ESEA), also known as the No Child Left Behind Act (NCLB), requires standards-based accountability for school districts and schools receiving Title I funds. One major component of this accountability policy is to report whether the districts and schools are making "adequate yearly progress" (AYP) based on their performance goals. Previous studies pointed out that some critical problems with present AYP-related efforts foreshadow technical challenges that lie 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, no attention has been paid to the uniqueness of rural schools. The following questions are raised particularly for rural schools: What are the implications of NCLB's performance-based accountability policy mandate for rural school improvement? How can we overcome technical measurement challenges that small, rural schools face in response to the policy mandate?

In this paper, I attempt to answer several questions through simulation analyses of data: how the NCLB's AYP formula would have worked if we had applied it to past school and student performance data, and what would have happened if we had applied methods different from those the current formula depends upon. First, is the AYP goal attainable and realistic? Second, is the AYP measure valid? Third, is the AYP measure reliable? Finally, is the AYP formula fair? All of these questions are examined with a focus on rural schools and their students in the field of mathematics. Using the National Assessment of Educational Progress (NAEP) from 35 states and state student assessment results from Maine, each of these four questions are addressed in the following sections. The answer to questions of who might win or lose from the current AYP process and how we can make this measurement strategy more feasible, valid, reliable, and fair for all may provide insight to guide policymaking.

There are some caveats in interpreting the findings of this study. The study focused on 8th grade mathematics achievement using national and state assessment data collected during the 1990s; that raises the question of whether or not the study findings can be generalized to different grades,subject areas, states, and time periods. As the 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 is different than what states have faced so far. Indeed, projection of future trends based on the past performance results might be wrong due to underestimating schools' potential progress expected under this new legislation. The results may have been different if schools had faced in the past 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, this paper can help policymakers and administrators become more aware of potential biases and pitfalls in evaluating academic progress made by rural schools and their students.

Is AYP Feasible?

Since the passage of the NCLB, much concern has been raised about what critics call 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 their AYP quota. It was estimated that up to 80 percent of schools in the states could be targeted as needing improvement or corrective action in the first few years (Olson, 2002, April 3). However, this kind of prediction needs an empirical verification, and the feasibility or attainability of the given AYP goal may vary significantly across the nation depending on which states and locations we look at.

NAEP is regarded as the nation's report card of student achievement in key subject areas. NAEP has changed its reporting criteria for type of location since 1990. In 1990, NAEP reported students' math proficiency by a type of community variable (advantaged urban, disadvantaged urban, and extreme rural) that combined community size with a school-level socioeconomic indicator. Discontinuing the classification due to the problematic nature of the variable, NAEP started reporting results by Census-based type of location since 1992.

The type of location classification system used in the 1992 and 1996 NAEP is based on geographic characteristics of the schools' locations and is related 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. This definition differs from the *Extreme Rural* category in past NAEP reports that encompasses students in nonmetropolitan areas with a population below 10,000 and where many parents are farmers or farm workers. In this section, schools in Central City, Urban Fringe, or Large Town are classified as "nonrural," and schools in Rural or Small Town as "rural."

In the 2000 NAEP, the same type of location variable was used, but it was not comparable with the results from previous years. This was due to the fact that NCES used a new method to identify the type of locations 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). As the NAEP data provides inconsistent information on rural vs. nonrural mathematics achievement because of changes in its definition of type of location, the only time period during the 1990s that allows us to examine rural students' progress in math achievement consistently for the nation and states is between 1992 and 1996. During this period, the most significant improvement occurred in rural schools nationwide. In 1996 rural students started to outperform non-rural students on the NAEP 8th grade mathematics assessment. Rural students' average math scale score was 276, whereas nonrural students' average score was 268; the 8-point gap amounts to approximately one-fourth of the pooled standard deviation. Specifically, students in Rural/Small Town scored 16 points more than students in Central City and 2 points more than students in Urban fringe/Large town on the 1996 8th grade math assessment.

The percentage of rural students at or above the NAEP Proficient Level increased from 17 to 25 across the nation between 1992 and 1996 in 8th grade mathematics (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 percent) every four years after 1996, we can project that the number of rural students at or above Proficient will rise to 53 percent by 2014. This figure remains far away from the goal of 100 percent Proficient. Even with the same amount of continuous gain since this time, it might take another 24 years to reach the 100 percent 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 lower achievement level Basic instead of Proficient as its target, the deficit would be smaller, but some of nonrural students might remain below the goal in 2014.

Insert Figure 1 about here

Despite these national trends, the mathematics achievement *levels* and mathematics achievement *gains* of rural students vary substantially (Lee & McIntire, 2001; Lee, 2002). First, some of the rural states performed at the top (e.g., Iowa and Maine), while others performed below the national average (e.g., Arkansas and Mississippi). Secondly, there are also interstate variations in rural students' mathematics achievement gain over the 1992-96 period (see Figure 2). Among the 35 states participating in the 1992 and 1996 NAEP 8th 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, West Virginia). For those significant states, the size of their gain scores are small to moderate, ranging from 4 to 12 points (approximately .1 to .4 in standard deviation units).

Insert Figure 2 about here

The percentage of students meeting the NAEP Proficient level of achievement also shows similar, uneven progress among different states. For example, Maine, one of the highest performing and the most improving states with a majority of students living in rural areas, shows a modest increase in the percentage of rural students at or above Proficient: from 25 to 30 between 1992 and 1996. Assuming a 5 percent gain for every four-year period, it is projected that Maine will have about 47.5 percent of its rural students meeting the Proficient level of achievement by 2014. These data tell us that the current goal of AYP imposed by the federal government is highly unrealistic, even for a state that is far ahead of others on the learning curve. When we lower our target to the Basic Level, the goal might become attainable: the projected estimate of the percentage of rural students in Maine at or above Basic in 2014 will be 98.5.

Is AYP Valid?

The NCLB Act requires each state to participate in biennial state assessments of 4th and 8th grade reading and mathematics under the National Assessment of Educational Progress (NAEP). Similarly, the NCLB Act requires each Local Education Agency (LEA) to participate, if selected, in the State NAEP. The law does not explicitly require that results from the NAEP be used as evidence to confirm progress on state tests, but its mandate that all 50 states now take part in the National Assessment of Educational Progress 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 comparison of the NAEP with state assessment results showed significant discrepancies in the size of statewide achievement gains(Lee & McIntire, 2002). This problem may apply to both rural and nonrural schools, although it is likely that rural and nonrural schools may respond quite differently to state curriculum and assessment mandates. Their different alignment may result in different gains on state assessments compared with national assessments.

Progress as reported by states is usually greater than progress as shown by the NAEP.

Table 1 compares Maine student performance improvement levels based on the NAEP with Maine Educational Assessment (MEA) 8th grade math assessment results. Because NAEP and MEA scores employ different scales, a common metric in standard deviation units was established. Specifically, student standard deviations as obtained from the MEA 1996 mathematics assessment results were used to compute MEA standardized gain, while Maine's standard deviations from the 1996 NAEP state assessment results were used to compute NAEP standardized gain. Table 2 breaks down these four-year achievement gains by type of school location. It shows that both rural and nonrural students made larger gains on state assessment than on the NAEP.

Insert Table 1 about here

Insert Table 2 about here

As shown in Tables 1 and 2, we find overall statewide academic improvement in Maine between 1992 and 1996 as measured by the MEA and NAEP. Score gains measured by the state assessment, however, are greater than gains observed in national assessment results (NAEP). Compared with NAEP, the MEA gains are approximately 2 times larger for all students, approximately 2 times larger for rural students, and 1.25 times larger for nonrural students. This information does not tell us which assessment is more accurate and valid, but shows that choosing a particular assessment can make a difference.

It is recommended that we use multiple evaluation measures if the measures we use lead to consequences for students or their school systems (see AERA, APA, & NCME, 1999). This is also true for evaluation of school AYP. There are possible hazards of evaluating school achievement gains based on a single measure. Using more than a single measure (e.g., NAEP, state assessment, district standardized test, and school/classroom assessment) may allow us to get a more comprehensive and balanced picture of student achievement and enhance the validity and fairness of evaluation.

Indeed, states would have significant flexibility in meeting new federal testing requirements, under draft regulations released by the Department of Education (Olson, 2002, March 6). States could use a combination of state and local assessments. States likewise could use either tests designed to assess students' achievement of state standards, or tests designed to measure achievement against national norms. States using nationally norm-referenced tests, however, would have to alter them to reflect fully the states' standards. If states also want to include local assessments in the state testing system, they have to ensure that local assessments were aligned with state standards and were of acceptable technical quality (Olson, 2002, January 9). Despite their flexibility, these testing requirements present more challenges to rural schools that often lack human and financial resources.

Is AYP Reliable?

Are school achievement gains as estimated through a successive cohort comparison method reliable? The current measure of AYP is based on comparison of successive student groups' performance at the same grade level and can be highly unreliable (Lee & Coladarci, 2002; Linn & Huag, 2002). This lack of reliability is potentially problematic for small, rural schools when the law requires reporting the progress of every major demographic subgroup in each school. The problem is that the smaller the sample size, the more unreliable are the measures based on information from that sample. This fact means first, that small schools will produce inherently unreliable scores and, second, that demographic cohorts withinin small schools will yield *very* small sample sizes and *very* unreliable scores. Inequalities among schools may result from the use of unreliable AYP and other accountability measures for the allocation of resources or federal aid.

It has been argued that states should switch from this successive cohort model assessing different cohort groups' achievement to a value-added model assessing the same cohort's achievement gains (Wheat, 2000). According to this alternative accountability model, 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 4th grade passing rates achieved by the same cohort of students. Although this approach can help cope with many problems associated with current AYP measures, it raises new technical difficulties and challenges by requesting that states track individual students' academic performance over time. The challenges include assessing students with the same or comparable tests as they move from one grade to the next grade and setting comparably rigorous performance standards across those different grades to determine any change in student proficiency level.

An analysis undertaken for this report shows that the successive cohort comparison is highly unreliable as a measure of academic progress in small rural schools. The state department of education web site usually provides publicly available information on school average performance and progress that can be used as official measures of AYP. 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 made publicly available by the National Center for Education Statistics (NCES). This combination of data allows one to compare academic progress in rural vs. nonrural schools in particular states.

The following analysis of Maine schools' 8th grade math achievement gains using the MEA data collected during the 1990-98 period, combined with CCD data for 2000. The analysis shows that, in small, rural schools, the estimation of academic progress 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 1990-98 period, which makes it hard to detect its overall performance trend. In contrast, the nonrural school shows a high level of stability with its generally upward performance trend. The difference in performance trends reflects their differences in school size and enrollment trends. The rural school had fewer than 40 students tested during the period under study, and the number changed substantially from year to year: 12 in 1995, 37 in 1996, 22 in 1997, and 37 in 1998. The nonrural school has larger than 200 students tested and the number was relatively stable: 245 in 1995, 241 in 1996, 254 in 1997, and 255 in 1998. These differences lead to the conclusion that the current AYP formula and the allocation of resources to schools based on such unreliable AYP measures can be misleading.

Insert Figure 3 about here

Under the NCLB AYP provisions, schools have the option of "using a uniform averaging procedure, 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. This process works in a school's favor when test scores decline, but it works against a school when scores rise. In order to see how well this provision can reduce variations in school performance trend, analyses of all Maine schools' 1990-98 MEA data were conducted with and without this rolling average procedure. Figure 4 shows the distributions of rural vs. nonrural schools' score variability as measured by the standard deviations of their 9-year scores. This comparison tells us that rural schools are much more unstable than their nonrural counterparts and that the use of rolling average procedure can help reduce the instability to a greater extent for rural schools.

Insert Figure 4 about here

It is particularly challenging to measure achievement gains made by even smaller demographic subgroups (racial and ethnic minorities, students with Limited English Proficiency [LEP], etc.) in small, rural schools. Statutory language does make provisions to exempt states from the requirement to report or use disaggregated data "in a case in which 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."

The U.S. Department of Education, however, does not provide guidelines to "determine and justify in its State plan the minimum number of students sufficient to yield statistically reliable information…" (MacQuarrie, 2002, September).

Is AYP Fair?

The current AYP policy targets its support and accountability (i.e., aid 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. It sets the same target for all schools with the same timeline, ignoring the substantial variations among schools in the demographic composition of their students. Studies demonstrate differences among urban, suburban, and rural schools in their students' academic achievement and also show that the variation in average school performance is closely related to many factors beyond schools' control (Lee, 2002; Lippman, Burns, & McArthur, 1996). Consequently, schools that happen to have many disadvantaged, low-achieving students like Title I participants and thus perform relatively poorly may be unduly penalized for failing to meet AYP targets.

Table 3 shows the 1996 NAEP math performance of 8th grade students participating in Title I programs and services in rural and nonrural schools nationwide. In both rural and nonrural schools, students who participate in Title I programs scored lower than their non-participating classmates. However, the achievement gap between Title I and non-Title I students is much more significant in nonrural schools than in rural schools. More importantly, Title I students in rural schools perform significantly better than their counterparts in nonrural schools, while non-Title I students in rural schools perform about the same as their counterparts in nonrural schools. This indicates that rural Title I schools might show better AYP performance than nonrural Title I schools.

Insert Table 3 about here

By setting uniform AYP target for every school, the current formula does not consider the influence of schools' initial performance status on their chance to meet the target. Higher-performing schools that are above the AYP target at the beginning will be able to meet the target much more easily than lower-performing schools that are initially below the target. Figure 5 illustrates two hypothetical schools, A and B, both of which are assumed to ultimately reach the goal of 100% proficient in 12 years, but have taken quite different paths. In this scenario, school A would meet the target throughout the 12-year period, whereas school B would never meet the target except year 12. This situation is ironic because school B made much greater progress than school A throughout the period. In order to avoid becoming a failing school consecutively, school B would have to increase its performance substantially during the first year to reach its target. Nevertheless, this large initial increase in performance is very unlikely to happen, considering how much time and energy it might take to break through the natural tendency of incremental change and to fully implement new programs for an effect.

Insert Figure 5 about here

Even among lower-performing schools, schools that are closer to the target initially are in a better position to meet the AYP target continuously. It may make more sense to set different targets for schools predicated upon the school's baseline status; every school would have its own AYP target, although the eventual goal would remain the same for all. Schools that initially performed at a lower level would be assigned a task of making relatively large gains – that is, meeting a higher AYP threshold – while initially higher performing schools would have to meet a relatively lower AYP threshold. Expected growth trajectories would be individualized, following the examples shown in Figure 5.

To test how different AYP-setting approaches work, I analyzed all Maine schools' 1990-98 MEA 8th grade math achievement data. I then compared results from using uniform AYP targets (the current formula) with results from setting individual AYP targets. First of all, the current AYP formulas were used to determine baseline and annual AYP targets in Maine. The average score at the state's 20th percentile school was treated as baseline (250) in 1990; hypothetical AYP targets were set above that baseline in increments of 10 points per year, so that the AYP target became 370 in 2002 – that is, 12 years from 1990. This hypothetical AYP target was a reasonably high standard considering the fact that 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 1990-98 period (see Figure 6).

Insert Figure 6 about here

Table 4 shows how many times rural and non-rural schools in Maine would have met the AYP target throughout the 1990-98 period under this "uniform" AYP formula scenario. Schools that were initially below the 20th percentile would have met the AYP target only 4 times out of 9 years, while schools that were initially above the 20th percentile would have met the target about twice as many times: both rural and nonrural schools followed similar patterns. Table 5 shows the results of applying the alternative, "individualized" AYP formula to the same data. Now schools that were initially below the 20th percentile would have met the AYP target 5 times, while schools that were initially above the 20th percentile met the target 6 times over the 9-year period. Therefore, this alternative AYP-setting approach tends to produce more equitable results.

Insert Table 4 about here

Insert Table 5 about here

This individualized AYP formula may be further adjusted to take into account possible significant changes in the composition of student body and the amount of resources available to schools over time. If a school takes substantially more disadvantaged minority students this year than the previous year, it might deserve a smaller quota of demonstrated progress. If a school gets a substantially larger amount of money for its academic improvement this year compared to the previous year, it might be reasonably assigned a greater AYP quota than before. Tying the AYP target to demographic and resource changes should produce fairer results by taking into account factors that are often unique to individual schools and beyond the control of schools.

Conclusion

The most imminent challenge to the measurement of rural schools' and their students' achievement gains comes from the Adequate Yearly Progress (AYP) mandate from the No Child Left Behind Act (NCLB). 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 racial and socioeconomic groups of students in rural schools narrowing? Rural educational policymakers and practitioners will face such questions, which 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 fewer resources, can provide "a model of strength" worth studying and emulating (Lee, 2002). The NCLB grants 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 the personnel and resources needed to compete (NECEPL, 2002).

Many poor small rural schools, however, are potentially threatened by unrealistically rigorous AYP targets and accountability measures (i.e., corrective actions and sanctions). The bottom line is that many disadvantaged schools in the nation and states are highly unlikely to reach the current AYP goal unless we lower the target achievement level or extend the timeline to reach the level. Rural schools may appear to have a better prospect of progress than their nonrural counterparts based on the past NAEP results, but it remains to be seen whether they can sustain past progress. Enormous variations across the nation in the academic status and progress of rural schools complicate this prediction (Lee & McIntire, 2001; Lee, 2002).

Many unresolved technical issues regarding the properties of achievement measures for evaluating school progress make the current approach to measuring AYP problematic. Any technical weaknesses and flaws of the measures might lead people to view the AYP as a political construct rather than a scientific indicator of school progress. The validity, reliability, and fairness issues are interrelated and, thus, should be tackled together. How can we make the AYP measure more technically sound? How can we make the AYP more valid, reliable and fair within the parameters of the current legislation and regulation?

For enhancing the validity and fairness of AYP, schools should be allowed to use multiple measures, including school and classroom assessments, to demonstrate their progress. State or local education agencies could use state assessment results to cross-check individual schools' self-reported AYP measures, while making sure that all of the assessments used for reporting AYP are aligned with the state's common curriculum standards and are of acceptable quality. Although this process might work under the current regulation, it might be very challenging for small rural schools that do not have adequate resources and staff to develop multiple assessment tools and prove their quality.

To enhance the reliability of AYP, schools should be allowed to combine their achievement measures from multiple years by using rolling average procedure that helps stabilize variations in their performance measures due to changes in student population and other confounding factors. Although this procedure is available under the current law, it remains difficult for rural schools to demonstrate consistent progress for their relatively smaller body of students and subgroups. Under the current AYP formula, small rural schools that are poorer and more vulnerable to environmental changes might risk failing to meet the AYP target and, therefore, be closed or consolidated. To make the AYP fair, the current law should be revised to give schools greater flexibility in setting pathways to their ultimate AYP goal (100% proficient). Individualized AYP targets could be set for each school according to its baseline performance level, with targets adjusted over time as a school's conditions change.

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., et al. (2001). The nation's report card mathematics 2000. Washington, DC: NCES (NCES Report No. 2001–517).

Hill, R. (1997). Calculating and reducing errors associated with the evaluation of adequate yearly progress. Paper presented at the annual assessment conference of the CCSSO (ERIC Publication No. ED 414307).

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. Greenwitch, CT: Information Age Publishing.

Lee, J. (2001). Interstate variations in rural student achievement and schooling conditions. ERIC Clearinghouse on Rural Education and Small Schools. ERIC Digest. Available at 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. Available at 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. Available at 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.

Lippman, L., Burns, S., & McArthur, E. (1996). Urban schools: The challenge of location and poverty. Washington, D.C.: National Center for Education Statistics.

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.

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.

Olson, L. (2002, January 9). Testing systems in most states not ESEA-ready. Education Week. Available at http://www.edweek.org/ew/newstory.cfm?slug=16test.h21.

Olson, L. (2002, March 6). Testing rules would grant states leeway. Education Week. Available at http://www.edweek.org/ew/ew_printstory.cfm?slug=25rules.h21.

Olson, L. (2002, March 13). Want to confirm state test scores? It's complex, but NAEP can do it. Education Week. Available at http://www.edweek.org/ew/ew_printstory.cfm?slug=26naep.h21.

Olson, L. (2002, April 18). 'Inadequate' yearly gains are predicted. Education Week. Available at http://www.edweek.org/ew/ew_printstory.cfm?slug=29ayp.h21.

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

Thum, Y. (2002). Design of School Performance and School Productivity Indicators: Measuring Student and School Progress with the California API. Working draft.

Wheat, D. (2000). Value-added accountability: A systems solution to the school accreditation problem. Thomas Jefferson Institute for Public Policy report. Available at http://www.thomasjeffersoninst.org/pdfs/VAA_2nd_edition.pdf.

Table 1
MEA and Maine NAEP Eighth Grade Average Math Scores, 1992 and 1996

Assessment	1992	1996	Raw Gain	Standardized Gain
MEA	305	350	45*	0.34
NAEP	279	284	5*	0.16

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

Table 2

MEA and Maine NAEP Eighth Grade Average Math Scores by Type of School Location, 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.

Table 3

Percentage of 8th Graders Participating in Title I Programs and Their Average NAEP Mathematics Scale Scores by Type of School Location

	Rural		Nonrural	
	Percentage	Average Score	Percentage	Average Score
Participated	9 (2.8)	263 (7.5)	15 (2.5)	238 (2.3)
Did not participate	91 (2.8)	278 (2.0)	85 (2.5)	273 (2.0)

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 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample.

Table 4

How many times Maine schools would have met AYP target during 1990-98 period under "Uniform-AYP" scenario by school performance level and location type

Performance Level	Locale	N	Average Number of Meeting AYP Target
Low	Nonrural	2	4.0
	Rural	29	4.4
High	Nonrural	23	8.3
	Rural	123	7.4

Note. Low-performing schools are the schools whose 8th grade math score was initially below the AYP target in baseline year 1990. High-performing schools are the schools whose 8th grade math score was initially at or above the AYP target in baseline year 1990.

Table 5

How many times Maine schools would have met AYP target during 1990-98 period under "Individualized-AYP" scenario by school performance level and location type

Performance Level	Locale	N	Average Number of Meeting AYP Target
Low	Nonrural	2	6.5
	Rural	29	6.2
High	Nonrural	23	5.3
	Rural	123	4.8

Note. Classification of low vs. high performing schools is the same as in Table 4.

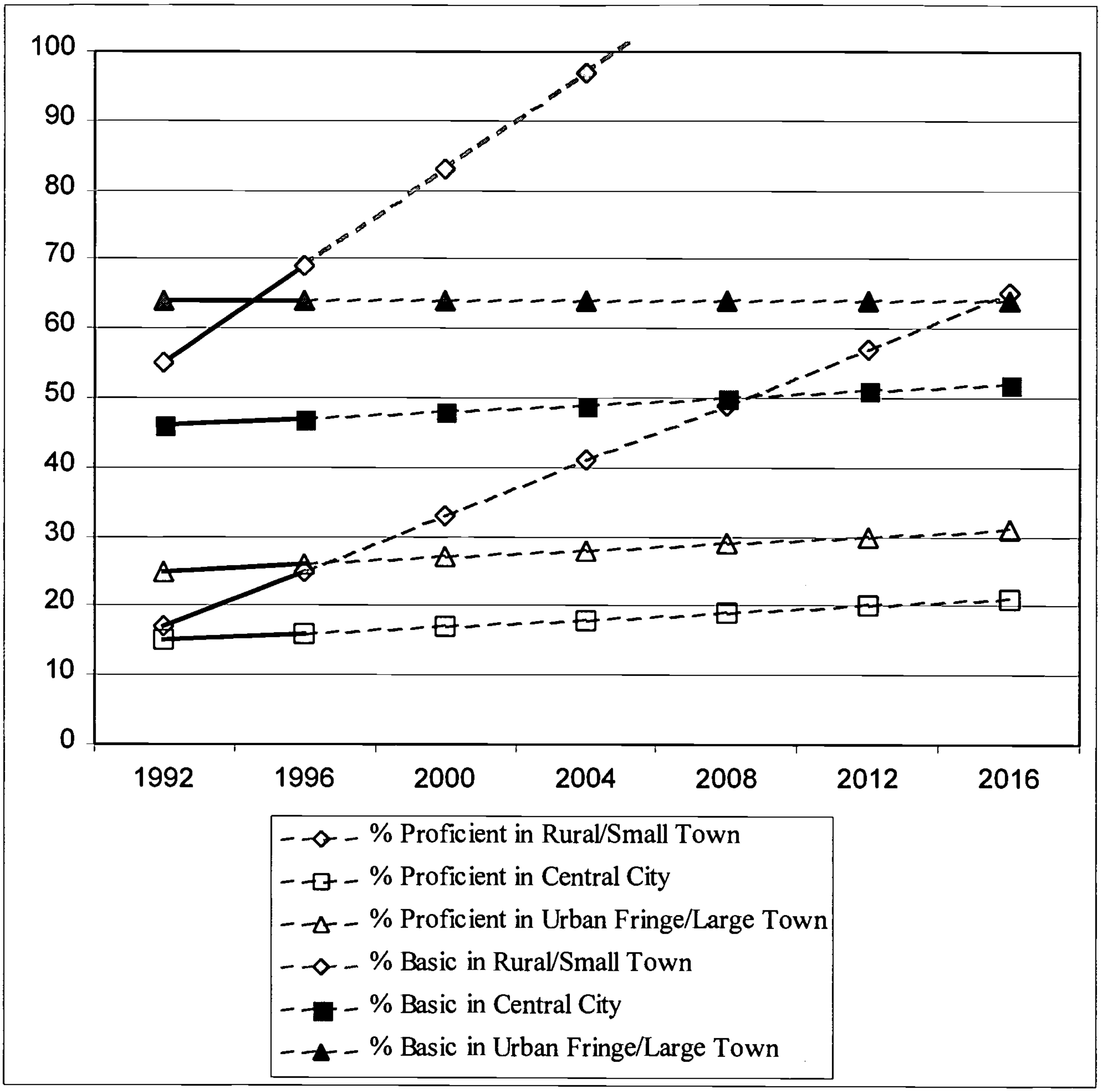

Figure 1. Trajectories of Percent Students At or Above Basic and Proficient Achievement Levels based on the 1992 and 1996 NAEP 8th Grade Math Assessment Results by Type of Location (actual measures are shown in solid lines and projected estimates are shown in broken lines)

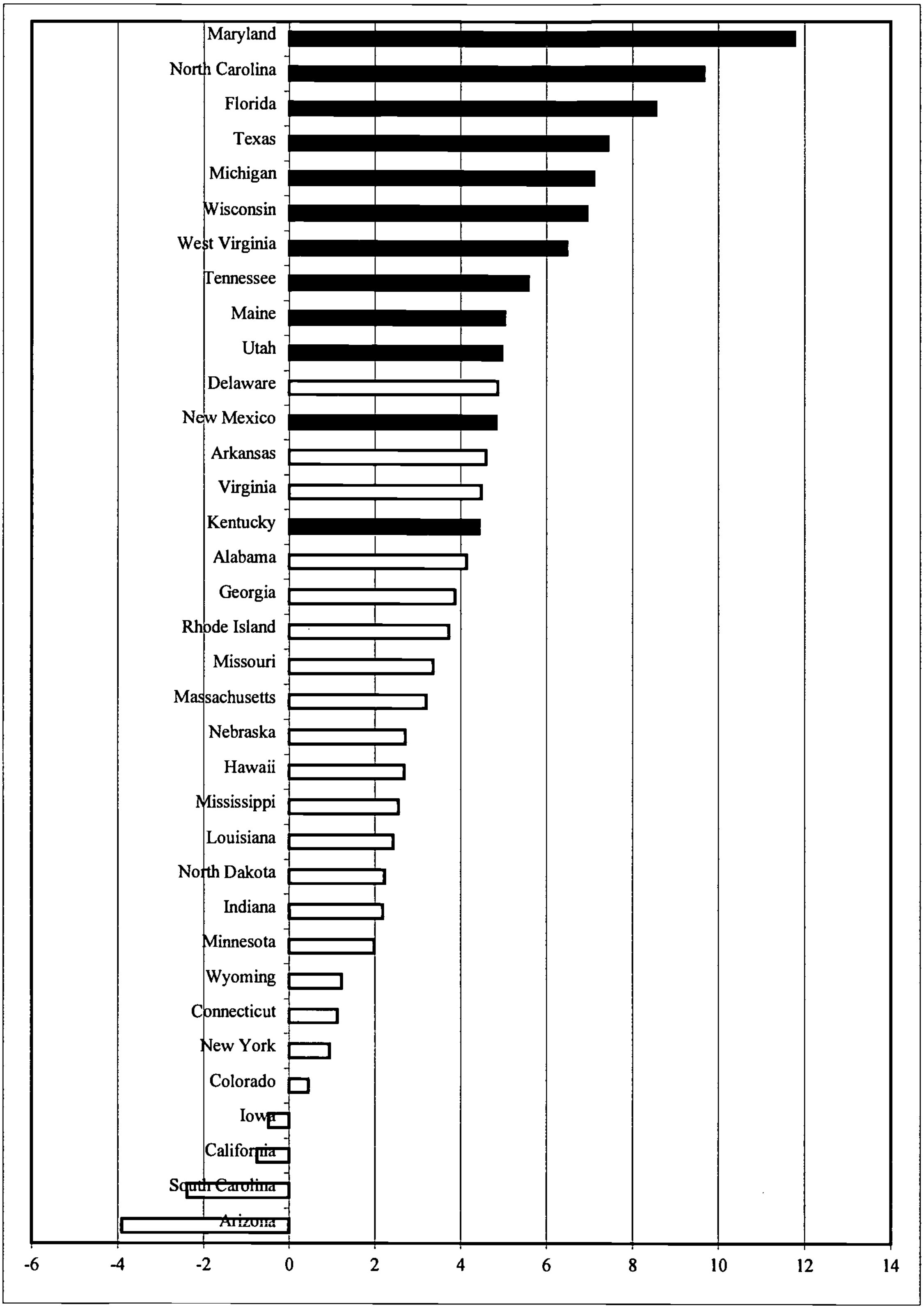

Figure 2. State average 8th grade NAEP math achievement gains from 1992 to 1996 in rural/small towns (N= 35 states; states with statistically significant gains are shown in black bars).

BEST COPY AVAILABLE

Figure 3. Two sample Maine schools' 8th grade MEA math performance trajectories

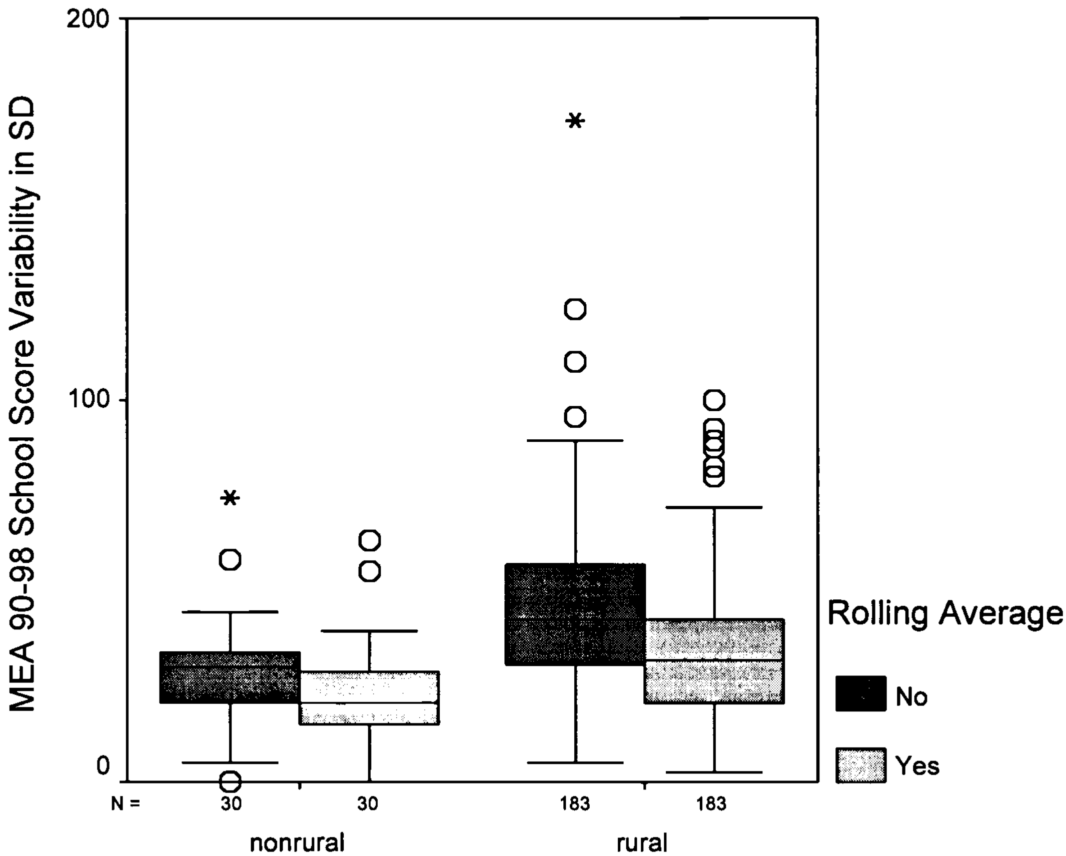

Figure 4. Rural and nonrural Maine schools' variability of 1990-98 MEA 8th grade math scores with and without rolling average procedure (N=213; 30 nonrural schools plus 183 rural schools)

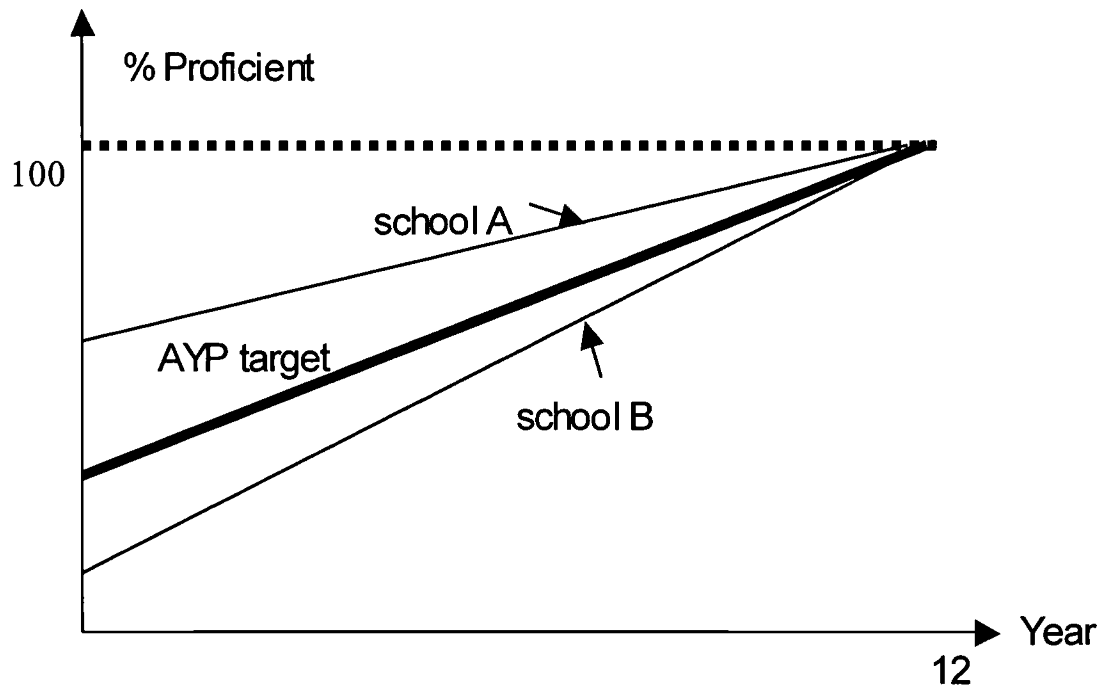

Figure 5. Hypothetical performance trajectories of two schools compared with their common AYP target line

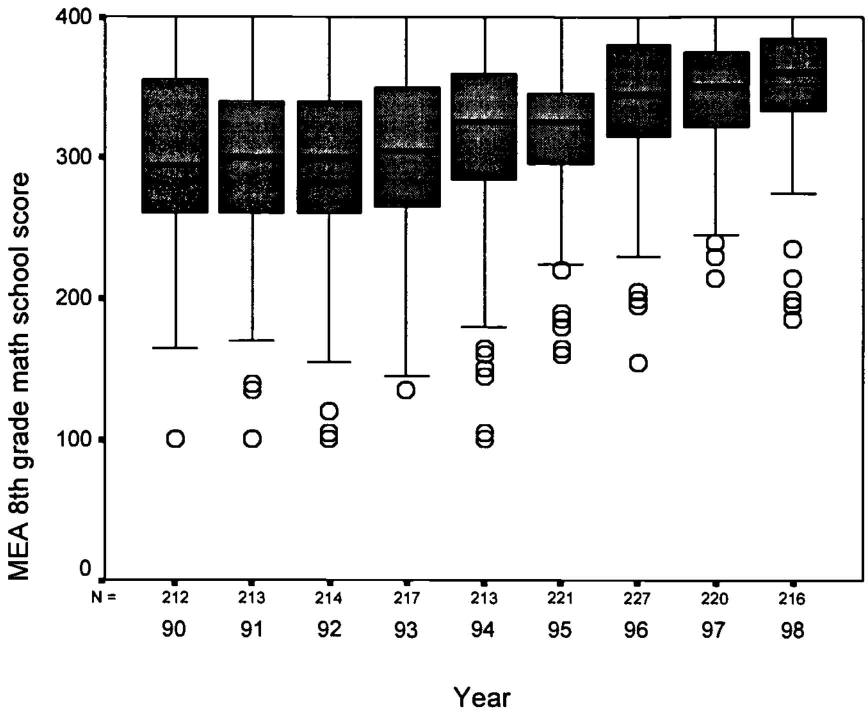

Figure 6. Distributions of Maine schools' MEA 8th grade math scores in 1990-98

NOTICE

Reproduction Basis

- [X] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

- [ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").