



POLICY BRIEF

The Devil is in the Details:

Rural-Sensitive Best Practices for Accountability Under No Child Left Behind

By Lorna Jimerson, Ed.D.

January 2004

RURAL TRUST POLICY BRIEF SERIES ON RURAL EDUCATION

Copyright © 2004 by the Rural School and Community Trust. All rights reserved.
Printed in the United States of America.

To order a copy of this publication, please contact the Rural School and Community Trust, or visit our website at www.ruraledu.org for a free downloadable copy (PDF file requires Adobe Acrobat Reader).

The Rural School and Community Trust
1825 K Street NW, Suite 703
Washington, DC 20006
Telephone: (202) 955-7177
Fax: (202) 955-7179
E-mail: info@ruraledu.org

The Rural School and Community Trust (Rural Trust) is the premier national nonprofit organization addressing the crucial relationship between good schools and thriving rural communities. Working in some of the poorest, most challenging rural places, the Rural Trust involves young people in learning linked to their communities, improves the quality of teaching and school leadership, advocates for appropriate state educational policies, and addresses the critical issue of funding for rural schools.

 Printed on recycled paper.

Table of Contents

Introduction	5
Rural-Sensitive No Child Left Behind Accountability	6
Rural-Sensitive State Accountability Plans	16
Alabama, Alaska and Arkansas	16
Louisiana, Maine and Mississippi	18
Montana, Nebraska and New Mexico	20
North Carolina, North Dakota, and South Carolina	22
Vermont, West Virginia, and Wyoming	24
Conclusions	26
Appendix A: What is a cell?	29
Appendix B: Selected Resources	31

Introduction

Across the country, states are concentrating their efforts on meeting the requirements and the spirit of the No Child Left Behind Act (NCLB). The implementation provisions and timelines are demanding and challenging for all districts, but are particularly daunting for rural and small districts. Although NCLB is quite prescriptive, the legislation does allow states some "wiggle room" in certain policy areas, many of which are especially important for rural and small schools. Some choices can help rural and small schools successfully navigate NCLB. Other choices may make the law more problematic.

This report focuses only on certain details of the accountability provisions of NCLB. We have identified 12 areas where state accountability plans can have special significance for rural schools and districts. We call policies that are beneficial to rural and small schools "Rural-Sensitive Best Practices." These policies most clearly recognize the realities and the challenges of schooling in rural areas. The report looks at 15 states with significant rural populations, (Alabama, Alaska, Arkansas, Louisiana, Maine, Mississippi, Montana, Nebraska, New Mexico, North Carolina, North Dakota, South Carolina, Vermont, West Virginia, and Wyoming) and examines the extent to which state accountability choices are "rural-sensitive" in each of these 12 areas.

Our intent here is not to "rate" states' accountability plans, but rather to illuminate some of the complexities of NCLB from a rural perspective. Though we have placed a checkmark (✓) next to state policies we believe are "rural-sensitive" in the tables that constitute the main portion of this report, we realize that a variety of factors led to some of these policy decisions; the lack of a checkmark does not indicate that the state is insensitive to rural issues. We understand that there are both competing state needs and competing NCLB requirements.

We wrote this report with two audiences in mind. First, the report is designed to help rural advocates isolate and understand some of the details of NCLB accountability that make a difference for rural schools

and districts. Second, we hope the report will be useful for state officials as they consider modifications to state accountability plans.

The report has three sections. The first section is a chart that examines each of the 12 policy areas and describes its significance for rural schools. For each area, we have identified the most "rural-sensitive" position. The second section examines how each of these 12 policy areas is treated in the NCLB plan for 15 of the most "rural" states. The last section summarizes the findings of our investigation, discusses other areas of importance, and offers reflections about NCLB accountability in rural places.

Rural-Sensitive No Child Left Behind Accountability

Accountability Element (in State NCLB Plans)	Significance and Rationale	Rural-Sensitive Best Practices
<p>1. Minimum cell size for AYP identification</p>	<p style="text-align: center;"><i>— Minimizing invalid AYP determinations because of small size —</i></p> <p>Small schools are especially vulnerable to being misidentified as failing Adequate Yearly Progress (AYP) because small numbers of students take the tests. Small numbers make test results statistically very weak. And since AYP judgments are made on the basis of these unreliable statistics, many small schools and districts probably will be incorrectly identified as "failing."</p> <p>The U.S. Department of Education recognized this problem and asked each state to identify a minimum cell size (or minimum N) for AYP determination. This number refers to the number of "data points" (or number of test-takers) used in each particular calculation for AYP. If the number of students taking the test is below this number, then states are not required to include this group in AYP calculations.</p> <p>Exactly, what is a "cell"? A cell is defined as each separate unit of analysis used in determining AYP. The number of cells varies by state, however. Some states have separate cells for each grade level that takes the state accountability tests (Mississippi, for example). Other states are combining test scores (for each subject) across all tested grades within each school. This does mean that test results are combined for different tests (i.e., different tests used in different grade levels). Though statistical validity of this procedure is questionable, combining data greatly reduces the number of cells used to determine whether a school meets AYP. See Appendix A for examples of cells in typical schools.</p> <p>The current range for minimum cell size in all states is from five to 200 (with some other statistical requirements). Though states recognize this problem, there is a lack of consensus about the most effective cell size for avoiding erroneous AYP judgments. However, in general, the larger the minimum N, the greater the protection for small schools against being misidentified as failing.</p>	<p>Large minimum N (40 or higher).</p> <p style="text-align: center;">or</p> <p>No minimum N set and use a confidence interval (see #5 below) to increase reliability for very small numbers.</p>

N = CELL SIZE

Accountability Element (in State NCLB Plans)	Significance and Rationale	Rural-Sensitive Best Practices
<p>2. Minimum cell size for reporting</p>	<p style="text-align: center;"><i>— Maintaining confidentiality in small schools —</i></p> <p>States must designate a minimum cell size for reporting purposes.</p> <p>States need to publicly report all assessment results for entire schools (and districts) and for each of the subgroups. The primary issue for small schools is confidentiality: when results are displayed by subgroups (like ethnicity and students with disabilities) individual students may inadvertently be easy to identify. Small schools and small communities are places where everyone is known and where maintaining confidentiality is more difficult.</p> <p>(Note: It can also be argued that if AYP identification is "unreliable" (e.g., $N < 40$), then reporting unreliable data is irresponsible and therefore the N for reporting should be at least as high as the N for AYP accountability. This is a reasonable position to take, though the focus here is the confidentiality, and an N of 20 should provide that.)</p> <p>In general, a higher minimum N for reporting provides the greatest protection of privacy for rural students and families.</p>	<p>Large N (20 or higher).</p>
<p>3. Minimum cell size for participation rate</p>	<p style="text-align: center;"><i>— Ensuring that absentees in a small school won't affect participation requirements —</i></p> <p>NCLB requires at least 95% participation in assessments for each subgroup. For very small schools, with small numbers of test-takers, one or two absent students can cause schools to fail AYP. For example, if two students are not tested in any particular subgroup of 20, then the school will fail AYP because of the participation rule.</p> <p>Not all states specifically address this issue, since there is not much flexibility in NCLB. A few states, however, have proposed other solutions to the "small N" problem.</p> <p>Some rural states have specified a minimum N for participation rates in order to account for very small numbers. Other states specify that the participation rate is met if N minus one ($N - 1$), or N minus two ($N - 2$) students take the tests. That is, regardless of the</p>	<p>Not required, but the most rural-sensitive position is to specify a minimum N.</p> <p>Again the higher the N, the better.</p> <p>$N > 40$ can be considered a high N.</p> <p style="text-align: center;">or</p>

Accountability Element (in State NCLB Plans)	Significance and Rationale	Rural-Sensitive Best Practices
(#3, cont.)	<p>percentile, if all students <u>except</u> one or two take the test, then the participation rate is met. In addition, at least one state (North Dakota) proposes using a confidence interval around participation rate. (See #5 below.)</p> <p>All of these alternative approaches are good practices and acknowledge the issues of very small numbers in rural schools.</p>	<p>Use a confidence interval around participation rate.</p> <p>or</p> <p>Allow N - 1 or N - 2 participation rate to meet AYP requirements.</p>
4. Separate cell size for disability	<p style="text-align: center;"><i>— Accounting for disproportionate numbers of students with disabilities in some small schools and districts —</i></p> <p>NCLB does not require a separate minimum N for students with disabilities, however, some states have included this in their plans.</p> <p>There are four main concerns about inclusion of "students with disabilities" in AYP calculations and disaggregating this data in public reports: (1) The incidence and severity of disabilities varies greatly from year to year and are beyond the control of the school. (2) Similarly, definitions and identification of disabilities are federally determined and not under local control. (3) There are some geographic locations that attract greater numbers of students with special needs due to access (or lack thereof) to specialized services. (4) Maintaining confidentiality is a concern in the public reporting of these students' performance.</p> <p>In very small schools and very small communities, all these issues become more pronounced and important.</p>	<p>The most rural-sensitive position is to specify a minimum N.</p> <p>Very few states do this. Almost any minimum number is more rural-sensitive than no minimum N. (Nebraska uses an N = 45.)</p>

Accountability Element (in State NCLB Plans)	Significance and Rationale	Rural-Sensitive Best Practices
<p>5. Use of a confidence interval (CI)</p>	<p style="text-align: center;"><i>— Providing for a “margin of error” in testing students in small schools for AYP —</i></p> <p>Determinations about meeting or failing AYP should be done in ways that minimize errors of misidentification. Sampling variation from year to year contributes to possible inaccurate judgments about school performance. Confidence Intervals (CI) are statistical treatments that can account for some of these problems.</p> <p>CIs therefore designate a "range" of scores that represent how a school is doing, rather than try to specify a precise score for the school's (or a subgroup's) performance. For example, if the calculated CI was $\pm 10\%$, instead of saying that exactly 50% of students in a school meet or exceed standards in math, a CI approach would determine that from 40% to 60% of students in that school would meet the standard 95 out of 100 times they took the test. A higher "level of confidence," for example 99% instead of 95%, results in a wider range. Also, a smaller number of data points (test-takers) results in a wider range.</p> <p>Most commonly, states are using CIs along with designating a minimum N. A few states (Montana and North Dakota) only use CIs and no minimum N, since the statistical formulas account for very small numbers.</p> <p>Note: A few states are using the CI model for other purposes such as calculating participation rates, "other academic indicator," and the "safe harbor" provisions for subgroup performance. Also, one state (Vermont) constructs a CI around the statewide proficiency targets, rather than around each school's results. Any of these uses can be considered "best practices" for small and rural schools.</p>	<p>Using a confidence interval is the most rural-sensitive position.</p> <p>A 99% CI is probably more favorable than a 95% CI, though there is a lack of consensus about this among psychometricians.</p>

Accountability Element (in State NCLB Plans)	Significance and Rationale	Rural-Sensitive Best Practices
6. Type of assessment(s)	<p style="text-align: center;"><i>— Creating a testing program that best measures student learning —</i></p> <p>No assessment (test) is perfect. The problems of multiple-choice tests have been noted for years. And with norm-referenced tests (NRT), the "norming" process forces 50% of students to always be below average. Criteria-referenced tests (CRT), usually constructed around standards-based curriculum, allow all students to potentially be "proficient." Most states are currently developing and/or using these CRT. CRT often have extended response or constructed response items. These tests are very expensive to develop and subject to scoring variations and errors.</p> <p><i>Multiple measures</i> are better indicators of student learning than any one assessment tool. Some state assessment plans use indices that combine results from a variety of assessment tools.</p> <p>Also, state testing programs fail to recognize local decisions about what to measure and how to measure student success. At least one state (Nebraska) is attempting to maintain local control of assessment in their state plan. Nebraska allows local districts to develop their own assessment plans, though they must meet six criteria established by the state and include math, language arts and the statewide writing assessment.</p>	<p>A rural-sensitive state assessment program would:</p> <p>(1) Avoid norm-referenced tests for NCLB accountability purposes.</p> <p>(2) Use multiple measures for assessing student achievement.</p> <p>(3) Allow local input in making testing decisions.</p>
7. Averaging across years	<p style="text-align: center;"><i>— Minimizing the effect of random fluctuations in test scores —</i></p> <p>NCLB allows states to use one, two or three years of data (i.e., use of a "rolling" average is permitted up to three years). The use of data from multiple years will mitigate some (not all) of the random fluctuations for very small schools.</p> <p>Some states maximize this provision in NCLB and will calculate AYP based on the most current three years of data <i>or</i> the current year, whichever is higher. That is, state policy allows districts to calculate the data both ways and pick the method that is most favorable.</p>	<p>Use an average from the three most current years (i.e., a three-year rolling average).</p> <p style="text-align: center;">or</p> <p>Use the higher of (1) the current year, or (2) an average from the three most current years.</p>

Accountability Element (in State NCLB Plans)	Significance and Rationale	Rural-Sensitive Best Practices
<p>8. Allow credit for improvement</p>	<p style="text-align: center;"><i>— Recognizing significant improvement in academic achievement —</i></p> <p>NCLB emphasizes <i>absolute performance</i> over "growth" or "improvement."</p> <p>NCLB does have the "safe harbor" provision for <u>subgroups</u>, which is essentially a "growth" component. However, the provision is quite prescriptive. If a subgroup does not make an AYP target, then the school can still "make" AYP through this provision given enough evidence of improvement over the previous year. Specifically, "safe harbor" requires at least a 10% decrease in the percentage of students who do not meet proficiency, <i>and</i> the subgroup must meet participation rates, <i>and</i> meet performance goals in the "other academic indicator." These provisions do not apply to the all-student determination, only to the subgroups.</p> <p>Some states, however, have their own unique ways of calculating (and recognizing) improvement, in addition to absolute performance level. These "growth" allowances are often incorporated into state plans by use of an "index" system of accountability. In addition, some states give "credit" for improvement in determining acceptable performance in graduation rate and/or with the "other academic indicator." (For example, Wyoming set a target graduation rate of 80%, but any improvement over the previous year's rate will be considered as having met AYP.)</p> <p>Recognition of improvement is especially critical for schools/districts that serve large proportions of disadvantaged children—and many rural districts fall into this category.</p>	<p>Index systems that include "credit" for growth or improvement.</p> <p style="text-align: center;">or</p> <p>Accepting "growth" or "improvement" as meeting AYP for any category, including graduation rate and the "other academic indicator."</p>
<p>9. Use of separate subject area results in determining AYP</p>	<p style="text-align: center;"><i>— Increasing the reliability of making judgments about low performance in any one subject area —</i></p> <p>NCLB requires that certain sanctions be imposed when schools or districts fail to meet AYP for two consecutive years. Annual assessments for AYP purposes must include math and reading, in grades 3-8 and at least one high school grade. (Future years will need to include science assessment.) Math and reading testing results must be reported separately.</p>	<p>Require failure in the SAME subject for two consecutive years to be considered as failing to meet AYP.</p>

Accountability Element (in State NCLB Plans)	Significance and Rationale	Rural-Sensitive Best Practices
<p>(#9, cont.)</p>	<p>States have the option of defining "failing AYP" as poor performance in the <i>same</i> subject area for two consecutive years. Or, they could define this as poor performance in <i>either</i> subject area for two consecutive years.</p> <p>For example, consider this situation. In Year 1, a school fails AYP targets in math, but meets AYP goals in reading. In Year 2, the school performs well in math, but fails AYP in reading. Using a "same" subject requirement, the school is not identified as "in need of improvement" using NCLB definitions. However, if a state uses an "either" subject requirement, then this school would be identified as "in need of improvement" under NCLB.</p> <p>A "same" subject approach is the most rural-sensitive since it avoids some of the pitfalls of random yearly fluctuations.</p> <p>(Note: There is another complicating issue here. It is unclear if the "two consecutive year" requirement is met when schools fail the <u>proficiency target</u> in one year, and then the <u>participation rate</u> [in the same subject] the following year. Most state plans leave this ambiguous.)</p>	
<p>10. Increments for intermediate goals</p>	<p>— <i>Allowing schools and districts adequate time to adjust to new accountability rules</i> —</p> <p>NCLB requires establishing intermediate performance goals for each subject area based on assessment data. These intermediate goals must lead to 100% proficiency by 2013-2014 and must increase in at least three-year increments.</p> <p>States have defined this in various ways. Some states have 12 equal annual steps (goals). Others have only four steps (changing the performance target once every three years). Other states are using a "mixed" model, with several initial three-year increments, followed by annual equal steps.</p> <p>Multi-year increments allow schools and districts more flexibility in meeting annual performance goals, since they have up to three years to reach the desired target.</p>	<p>Three-year increment (the minimum required).</p>

Accountability Element (in State NCLB Plans)	Significance and Rationale	Rural-Sensitive Best Practices
(#10, cont.)	<p>Some educators are concerned that permitting smaller initial increases (i.e., multi-year increments), will create more difficulties for schools, since the <u>future</u> increases will need to be <u>larger</u>. In spite of this valid concern, we believe that use of three-year increments is the most rural-sensitive position since it more clearly recognizes the real challenges in improving learning for <i>all</i> children and allows schools, and districts the most time to adapt to new assessment systems, new curricula and new standards.</p>	
11. Grade-span of starting points (and subsequent performance targets)	<p style="text-align: center;"><i>— Minimizing the number of separate AYP judgment points for each school and district —</i></p> <p>NCLB stipulates precise formulas for calculating the "starting points" for establishing AYP performance targets in separate subject areas. States have leeway, however, in establishing different performance goals for (1) each grade tested; (2) for each grade-span; or (3) for all tested grades, K-12. That is, states can have one math "starting point" for grade 3 and another for grade 4, etc. Or, states can have one math starting point for grades 3-5, and another starting point for grades 6-9, etc. Or states can have one uniform math starting point for all grades tested, K-12.</p> <p>Currently, most states only have assessments in one grade per grade-span, with other grade-level tests "in development." For practical purposes at this time, the decision is whether to calculate starting points across ALL grades or to calculate starting points PER grade (or grade-span).</p> <p>Either approach has advantages and disadvantages. Grade-level (or grade-span) starting points may be seen as more advantageous since performance targets will be closer to the "actual" achievement levels for students in each grade, thus more "reality based." However, grade-level (or grade-span) starting points may <u>dramatically</u> increase the number of separate AYP "tests" that each school (and each district) must pass in order to meet AYP.</p> <p>For example, assume grade-level targets in 4th and 8th grade in a K-8 school. This scenario will result in separate cells for 4th and 8th grade math for the all-students group and for</p>	<p>Use an "all-grade" target for each subject.</p> <p>In states with grade-level (or grade-span) targets, schools and districts MUST be able to average student performance across grades within each school and district.</p>

Accountability Element (in State NCLB Plans)	Significance and Rationale	Rural-Sensitive Best Practices
#11, cont.)	<p><i>each</i> separate subgroup in both grades. Depending on the number of ethnic groups, this school (or district) may need to meet 74 different AYP determinations, not the needed 37 if there were only one math target for the entire school. (The situation is even more extreme for K-12 schools, where there might be three grade-span targets for each subject, calculated for each subgroup.)</p> <p>Separate grade-level starting points (and targets) increase the chances of "failing" in some category for schools with wider grade-spans, and for rural schools that often have K-8 or K-12 configurations, this is <u>very</u> problematic. Some states have tried to avoid this problem by averaging performance data across grade-levels within a school and by averaging targets. This is analogous to having one statewide target.</p> <p>In general, the fewer the number of separate "cells," the less likely that small schools and small districts will be mistakenly identified as failing AYP. Thus the most rural-sensitive position is to have a single all-grade starting point for each subject area. (See Appendix A for examples.)</p> <p>Alternatively, if a state has separate starting points, there should be some mechanism created for combining test results so that each school has a minimum number of AYP determinations (cells). And in all cases, the requirements for minimum cell size and/or a CI needs to apply at the level in which decisions are made about AYP.</p>	
12. Special assessment procedures for very small schools	<p style="text-align: center;">— <i>Ensuring that very small schools are also evaluated</i> —</p> <p>If states set a minimum cell size high enough to maintain statistical reliability, there is some probability that small schools that are having difficulties will be "under-identified." That is, the by-product of procedures to avoid over-identifying small schools might be that schools that <u>need</u> technical assistance are not identified either. Thus, some states are putting in special systems to assess academic achievement and progress in very small schools (e.g., Vermont).</p>	Review progress of each individual small school, using a variety of measures.

**Rural-Sensitive
State Accountability Plans**

**Rural-Sensitive State Accountability Plans*
for Alabama, Alaska and Arkansas**

Accountability Element	ALABAMA	ALASKA	ARKANSAS
1. Minimum cell size for AYP identification	N = 40; if N < 40 for subgroup won't count; if N < 40 for all-student group, will determine AYP with note that AYP is based on less than minimum required N of 40. ✓	N = 21 for subgroups. For all-school or all-district, will calculate AYP even if N ≤ 20.	N = 25
2. Minimum cell size for reporting	N = 10	N = 21 for subgroups. For all-school or all-district, will report assessment results if N ≥ 5. ✓	N = 10
3. Minimum cell size for participation rates	N = 40; if N < 40, participation rate is met if N - 2 students (number enrolled minus two) participate in assessment. ✓	Minimum N = 21 for subgroups. If 22 > N > 40, will meet participation rate if all but two students are tested (N - 2). If N > 41, will need to meet 95% rule. ✓	
4. Separate cell size for disability			
5. Use of confidence interval (CI)		CI of 99% applied around assessment results. ✓	CI of 95% applied around assessment results. ✓

* Blanks on the charts indicate that this area was not included in the state consolidated plan.

Accountability Element	ALABAMA	ALASKA	ARKANSAS
6. Type of assessment(s)	CRT (in development); also writing assessment in some grades. ✓	Use both CRT and NRT: CRT in grades 3, 6, 8 and NRT in grades 4, 5, 7, 9. Also have a high school Graduation Qualifying Exam in grade 10. ✓	CRT ✓
7. Averaging across years	Use higher of current year, or most recent three-year average. ✓	One year.	Use higher of current year, or most recent three-year rolling average. ✓
8. Allow credit for improvement	Improvement in attendance and graduation rate = AYP. ✓		
9. Use of separate subject area results in determining AYP	Same subject area for two consecutive years. ✓	Same subject area for two consecutive years. ✓	Low performance in either subject area for two consecutive years results in failure to make AYP.
10. Increments for intermediate goals	First three increments are for two-years. Remaining are annual steps.	Three-year increments. ✓	Twelve equal increments.
11. Grade span of starting points	Two starting points—for grades 3-8 and for grade 11.	All tested grades combined. ✓	Three starting points, by grade-span, K-5, 6-8, 9-12.
12. Special assessment procedures for very small schools			

Rural-Sensitive State Accountability Plans for Louisiana, Maine and Mississippi

Accountability Element	LOUISIANA	MAINE	MISSISSIPPI
1. Minimum cell size for AYP identification	N = 10, with CI of 99%.	N = 20	N = 40; (if $N \leq 40$, will still use results but include statement that "may not be reliable because of small number of students.") ✓
2. Minimum cell size for reporting	N = 10	N = 10; Won't report if $N > 10$, if subgroup reporting can reveal identification of students. If too small for reporting, will still determine AYP based on participation rate and other indicators.	N = 10
3. Minimum cell size for participation rates	N = 40 ✓	N = 41 ✓	
4. Separate cell size for disability			
5. Use of confidence interval (CI)	CI of 99% applied around assessment results. ✓	CI of 95% applied around assessment results. ✓	CI of 95% applied around assessment results. ✓
6. Type of assessment(s)	CRT and NRT. Developing augmented NRT with writing prompt. ✓	MEA (criteria-referenced) and local assessments. ✓	CRT ✓

N = CELL SIZE

Accountability Element	LOUISIANA	MAINE	MISSISSIPPI
7. Averaging across years	AYP is based on one year. State assessment system incorporates both AYP and total school growth, an index using two years of data.	Two-year rolling average.	One year.
8. Allow credit for improvement	Louisiana assessment system calculates total school growth. ✓		Growth index used in state assessment system. ✓
9. Use of separate subject area results in determining AYP	Need two consecutive years of low performance in either subject area.	Same subject area for two consecutive years. ✓	Same subject area for two consecutive years. ✓
10. Increments for intermediate goals	Mixed model: first three increments are at three-year intervals. Last three steps are yearly increments. Total of six equal increments. ✓	Accelerating curve. First two increments are for three-year periods, then annual steps after that. Initial increments are smaller. ✓	Three-year increments. ✓
11. Grade span of starting points	All tested grades combined. ✓	Three starting points, by grade level: grades 4, 8, and 11.	Seven starting points, by grade level: grades 3, 4, 5, 6, 7, 8, and 10.
12. Special assessment procedures for very small schools		If N < 20 for these two years, will use three-year average. ✓	

Rural-Sensitive State Accountability Plans for Montana, Nebraska and New Mexico

Accountability Element	MONTANA	NEBRASKA	NEW MEXICO
1. Minimum cell size for AYP identification	No set N, CI of 95%. ✓	N = 30	N = 25
2. Minimum cell size for reporting	N = 10	N = 10	N = 10
3. Minimum cell size for participation rates	N = 40; if N < 40 will consider as meeting participation rate goal, if participation = N - 2. ✓	No one separate testing "event." Participation calculated as number of students who participated in at least 75% of all assessments or who were tested on at least 75% of all standards. ✓	N = 25
4. Separate cell size for disability		N = 45 ✓	
5. Use of confidence interval (CI)	CI of 95% for setting minimum cell size (see #1 above). ✓	No single cut score. No CI necessary. (Locally developed assessment systems.)	
6. Type of assessment(s)	Current assessment is NRT. Plans to change assessment (add CRT items) by spring 2004.	Locally developed assessment systems. Contain CRT, NRT and writing assessment. Assessment portfolio needs to meet six criteria. Must include math and language arts (at	Use both CRT and NRT now. Moving to all CRT by 2004-2005 (in both English and Spanish). ✓

N = CELL SIZE

Accountability Element	MONTANA	NEBRASKA	NEW MEXICO
(#6, cont.)		present, in grades 4, 8, and 11). Also must include statewide writing assessment (grades 4, 8, and 11). ✓	
7. Averaging across years	Three-year rolling average. ✓	One year.	One year.
8. Allow credit for improvement	Credit for improvement for attendance and graduation rate. ✓	Credit for improvement in writing assessment and graduation rate. ✓	Higher than expected growth rate can get off failing AYP cycle. ✓
9. Use of separate subject area results in determining AYP	Same subject area for two consecutive years. ✓	Same subject area for two consecutive years. ✓	Same subject area for two consecutive years. ✓
10. Increments for intermediate goals	Undetermined yet. First increment in 2004-2005.	Three-year increments. ✓	Annual equal increments.
11. Grade span of starting points	One starting point for all tested grades combined. ✓	Three starting points for grades 4, 8, and 11.	Three starting points for grades 4, 8, and high school.
12. Special assessment procedures for very small schools		All schools are assessed. To be reviewed in future. May use multiple years of data for very small schools. ✓	If N < 10 for one grade level, will calculate AYP using three-year rolling average. ✓

Rural-Sensitive State Accountability Plans for North Carolina, North Dakota and South Carolina

Accountability Element	NORTH CAROLINA	NORTH DAKOTA	SOUTH CAROLINA
1. Minimum cell size for AYP identification	N = 40; (will still include schools if N < 40, but will flag results as "based on less than 40 students.") Also note that North Carolina's assessment system uses an N = 30. ✓	No set N. CI at 99%. ✓	N = 40 If N < 40, will average last three years for AYP if possible. ✓
2. Minimum cell size for reporting	N = 5	N = 10	N = 10
3. Minimum cell size for participation rates	N = 40		N = 40
4. Separate cell size for disability			
5. Use of confidence interval (CI)		No minimum N. Use CI of 99%. ✓	
6. Type of assessment(s)	CRT ✓	Using an augmented NRT with additional standards-based items.	PACT (Pametto Achievement Challenge Test). 75% multiple choice; 25% open-ended. (Standards based—CRT) ✓

N = CELL SIZE

Accountability Element	NORTH CAROLINA	NORTH DAKOTA	SOUTH CAROLINA
7. Averaging across years	Three-year rolling average. ✓	Three-year rolling average. ✓	Use higher of current year, or average of most recent three years. ✓
8. Allow credit for improvement	"Growth" incorporated into ABC system. Also both attendance and graduation rate meet goal if there is improvement. Growth must be .1% increase or more. ✓		
9. Use of separate subject area results in determining AYP	Same subject area for two consecutive years. ✓	Same subject area for two consecutive years.	Same subject area for two consecutive years. ✓
10. Increments for intermediate goals	Three-year increments. ✓	Three-year increments. ✓	Three-year increments. ✓
11. Grade span of starting points	Two starting points, by grade span: grades 3-8 and 10.	Three starting points, by grade level: grades 4, 8, and 12.	One starting point for all tested grades combined. ✓
12. Special assessment procedures for very small schools			

Rural-Sensitive State Accountability Plans for Vermont, West Virginia and Wyoming

Accountability Element	VERMONT	WEST VIRGINIA	WYOMING
1. Minimum cell size for AYP identification	N = 80 for two years (on average N = 40). ✓	N = 50 (transition year; using SAT-9 for 2002-2003 only). Plans to determine minimum N using CI's and new WESTEST assessments. ✓	N = 30; (if N < 30, will use Wyoming's "Body of Evidence" to review performance of all schools.)
2. Minimum cell size for reporting	N = 11	N = 10	N = 6; (if N < 6, will use two or three years of data to get to N = 6 or more.)
3. Minimum cell size for participation rates		N = 50 ✓	N = 40 ✓
4. Separate cell size for disability			
5. Use of confidence interval (CI)	CI of 99% applied around target goals. ✓	Plan to use CI beginning next year.	CI of 95% applied around assessment results. ✓
6. Type of assessment(s)	CRT (including grade 2 Developmental Reading Assessment—DRA). ✓	Present system is NRT, transitioning to CRT.	CRT ✓
7. Averaging across years	Two-year rolling average.	Plan to use "multiple years." Baseline was established using two-year average.	Two-year rolling average.

N = CELL SIZE

Accountability Element	VERMONT	WEST VIRGINIA	WYOMING
8. Allow credit for improvement	Improvement in DRA and graduation rate = AYP. ✓	Requirements to meet AYP include "growth" or "improvement" in SAT-9, attendance and graduation rate. ✓	Credit for improvement for graduation rate even if below target of 80%. ✓
9. Use of separate subject area results in determining AYP	Same subject area for two consecutive years. ✓	Same subject area for two consecutive years (or same indicator, such as graduation rate or attendance or participation). ✓	Same subject area for two consecutive years. ✓
10. Increments for intermediate goals	Three-year increments. ✓	Mixed model proposed. First increment in 2005-2006, annual increments after that.	Accelerating curve. Two-year increments. ✓
11. Grade span of starting points	Four starting points for grades 2, 4, 8, and 10, but have calculated combined starting points for grade-spans so that each school or district only has <u>one</u> proficiency target per subject. ✓	Two starting points proposed. One for grades K-8, other for grades 9-12.	Three starting points, by grade: grades 4, 8, and 11.
12. Special assessment procedures for very small schools	Small school review for ALL schools if N < 30. Small school review for schools that fail AYP and N is between 30 and 79. ✓	Small schools (N < 50) will use most current three-year average for AYP determination. ✓	If N < 30, will use Wyoming "Body of Evidence" to review performance of all schools. ✓

Conclusions

Variability. This snapshot of the 12 policy areas across 15 states presents a complex mosaic of states' efforts to implement NCLB. Some states are clearly attempting to make NCLB accountability "work" for small schools and districts (e.g., by using CI), while others seem less concerned. Some states are maximizing the act's available flexibility (e.g., accepting growth for graduation rate), and others are accepting a more conventional interpretation of the law. The variability of state responses is notable.

With NCLB still in its infancy, some of the implications of these state policy choices for rural schools and districts remain uncertain. This is the first year that schools "in need of improvement" were identified with the current law. As NCLB begins to play out, doubtless many other implementation issues will emerge. We believe, however, that states that have adopted more rural-sensitive best practices are best serving the needs of the 8 million rural students in our country.

NCLB accountability as a moving target. The story of how states implement NCLB is an evolving tale. Although all states had their "consolidated" plans (for accountability) approved, most states received a follow-up letter after approval from the U.S. Department of Education (USDOE) requesting more information and/or modifi-

cations. Some states have responded to this letter and others are still in the process of responding (these letters are available at www.ed.gov/admins/lead/account/letters.) Though most of the information used in this report came from the approved state plans, interviews with certain state officials indicated that some details were still being "negotiated" with the USDOE.

The USDOE has promised a series of regulations and guidance about specific aspects of the law. While many have already been released, other promised regulations are still unavailable, leaving states with some ambiguity about exact implementation requirements.

In addition, some state officials indicated that their own education departments were going to consider further changes in their state accountability plans. One official remarked that after reading what other states were doing, he had some changes he wanted to pursue. State plans can be amended under NCLB; however, all changes need to be approved by the USDOE.

Another aspect adding to the evolution and confusion is the inconsistency of the federal approval process. Some states have had elements approved by the USDOE that were denied to other states, apparently dependent on how states justified a par-

ticular element that they wanted. *Statewide Educational Accountability Under NCLB* published by the Council of Chief State School Officers (July 2003) documents the recent history of state approval and highlights these differences.

Future directions: Other issues and other states. There are other aspects of accountability that demand attention beyond the 12 discussed here that we omitted primarily for simplicity and so that we could focus on those elements that most clearly impact rural districts. Also, some of the relevant components are still being developed (and/or negotiated) in many states (such as the required reward systems).

Among the missing issues not considered in this report are two very crucial assessment requirements: for students with disabilities and for students with limited English proficiency. We believe that assessment of these two subgroups will present major obstacles for many schools and districts in rural, suburban, and urban areas and that they deserve separate investigation in their own right.

In addition, though this report only covers 15 states with significant rural populations, other states have proposed some interesting rural-sensitive strategies that warrant attention. For example, Texas and New York have complex rules about "minimum

Ns" that vary the N according to the total population of the school. Wisconsin has set a very high N for students with disabilities.

Lastly, although this report only focuses on AYP (accountability), it is important to recognize other areas of NCLB that will have major significance for rural schools and districts, including two of particular concern: (1) the requirements for highly qualified teachers (and para-educators) and (2) the financial burden of implementing NCLB.

The Rural Trust will be tracking all these issues as they develop, and issuing additional reports as appropriate.

Consequences of rural-insensitivity. NCLB is basically a suburban-urban law. It is insensitive to many of the needs and problems of rural schooling. It ignores the reality of rural places. It allows little room for the values of rural communities. It puts small schools in a very vulnerable place. In spite of this rural-insensitivity, states can craft their accountability plans to minimize harm for rural and small schools. Our hope is that in curtailing potential harm, potential benefits can emerge.

NCLB is crafted so there are many ways for schools to fail. And rural and small schools are particularly vulnerable to being mistakenly identified as failing, especially because of very small numbers, a fate that can and will be devastating for many rural districts. The required NCLB sanctions are severe, and costly. At the very least, being identified as failing AYP can lead to a loss of public confidence. At the very worst, rural and small schools and districts will be subject to takeovers, privatization and consolidation. With such high stakes, it is crucial that states do whatever they can to ensure that their state accountability plan is valid, fair and sensitive to rural communities

Appendix A: What is a cell?**Examples of the different cells used in determining Adequate Yearly Progress for a typical school**Example 1: Using one proficiency target across all grades

The example below assumes that there is one performance target (i.e., one starting point) for all students in the school for each subject area. That is, in a K-6 school, for example, all students who are tested (say in grades 3, 4, 5, and 6) will need to meet the same proficiency targets. This implies that all math scores, for example, will be combined into one math score across all tested grades.

This example also assumes that there are five racial/ethnic groups that meet the minimum N requirement. Many rural schools will not meet this requirement, though the district might.

Each "square" below represents a separate cell and must adhere to the minimum "cell size" as adopted by the state. And if the state is using a confidence interval, this should also be applied to each cell. This is a "typical" AYP cell matrix and contains 37 different cells—or 37 different AYP calculations.

	Reading/Language Arts		Mathematics		Other Academic Indicator
	Participation Rate	% Meeting Standard	Participation Rate	% Meeting Standard	
All-Students					
Economically Disadvantaged					n/a
Racial/Ethnic Group 1					n/a
Racial/Ethnic Group 2					n/a
Racial/Ethnic Group 3					n/a
Racial/Ethnic Group 4					n/a
Racial/Ethnic Group 15					n/a
Students with Disabilities					n/a
LEP Students					n/a

Example 2: Using separate proficiency targets for each grade

This example shows what happens if schools are required to calculate and report separate scores for each grade level. If states designate separate starting points for each grade, and ask schools to determine AYP based on these grade-level expectations, then the number of distinct cells is greatly increased. In this example, there now are 148 cells. Each of these must pass the state-determined minimum N size (and confidence interval if used). This greatly increases the chances of schools or districts failing at least some part of the AYP provisions.

	Reading/Language Arts								Mathematics								Other Academic Indicator			
	Participation Rates				% Meeting Standard				Participation Rates				% Meeting Standard							
	3rd	4th	5th	6th	3rd	4th	5th	6th	3rd	4th	5th	6th	3rd	4th	5th	6th	3rd	4th	5th	6th
All-Student																				
Economically Disadvantaged																	n/a	n/a	n/a	n/a
Racial/Ethnic Group 1																	n/a	n/a	n/a	n/a
Racial/Ethnic Group 2																	n/a	n/a	n/a	n/a
Racial/Ethnic Group 3																	n/a	n/a	n/a	n/a
Racial/Ethnic Group 4																	n/a	n/a	n/a	n/a
Racial/Ethnic Group 5																	n/a	n/a	n/a	n/a
Students with Disabilities																	n/a	n/a	n/a	n/a
LEP Students																	n/a	n/a	n/a	n/a

Appendix B: Selected Resources

The following organizations have websites with good information about accountability and/or rural issues and the No Child Left Behind Act:

American Association of School Administrators (AASA) - www.aasa.org

- *No Child Left Behind: A Guide for Small and Rural Districts* - (from AASA and National Association of State Boards of Education) www.aasa.org/government_relations/rural/NCLB_and_rural_schools.PDF

Council of Chief State School Officers (CCSSO) - www.ccsso.org

- State Plan Summaries - www.ccsso.org/content/pdfs/StatePlanSummaries.pdf
- *Making Valid and Reliable Decisions in Determining Adequate Yearly Progress* (December 2002) - www.ccsso.org/Federal_Programs/nclb/3241.cfm
- *Statewide Educational Accountability under NCLB—Central Issues Arising from an Examination of State Accountability Workbooks and the U.S. Department of Education Reviews Under the No Child Left Behind Act of 2001* (July 2003) - www.ccsso.org/content/pdfs/StatewideEducationalAccountabilityUnderNCLB.pdf

Education Commission of the States (ECS) - www.ecs.org

National Association of State Boards of Education (NASBE) - www.nasbe.org

National School Boards Association (NSBA) - www.nsba.org

National Education Association (NEA) - www.nea.org

(continued on next page)

The Rural School and Community Trust - www.ruraledu.org

- *Gallup Goes to School: The Importance of Confidence Intervals for Evaluating 'Adequate Yearly Progress' in Small Schools* by Ted Coladarci - <http://www.ruraledu.org/docs/nclb/coladarci.htm>
- *Special Challenges of the 'No Child left behind' Act for Rural Schools and Districts* by Lorna Jimerson - www.ruraledu.org/docs/nclb/jimerson.htm

U.S. Department of Education - www.ed.gov/nclb/

- State Accountability Plans - www.ed.gov/admins/lead/account/stateplans03/index.html

Also see:

Volatility of School Test Scores: Implications for Test-Based Accountability Systems by Thomas J. Kane and Douglas O. Staiger (August 2001) - www.dartmouth.edu/~dstaiger/Papers/KaneStaiger_brookings2002.pdf

The Rural School and Community Trust
1825 K Street, N.W. Suite 703
Washington, DC 20006
(202) 955-7177 phone
(202) 955-7179 fax
www.ruraledu.org

Helping Rural Schools and Communities Get Better Together