

**Effect of Minimum Cell Sizes and Confidence Interval Sizes for Special
Education Subgroups on School-Level AYP Determinations**

Mary Ann Simpson, Brian Gong, and Scott Marion

Center for Assessment

September 26, 2005

Executive Summary

This study addresses three questions:

- First, considering the full group of students and the special education subgroup, what is the likely effect of minimum cell size and confidence interval size on school-level AYP determinations?
- Second, what effects do the changing minimum cell sizes have on inclusion of special education students, especially for schools that are declared as “meeting AYP”?
- Third, with the NCLB requirement that schools assess grade levels 3-8 in their AYP calculations beginning in the 2005-2006 academic year, what is the likely effect of including these additional students in school-level AYP determinations?

To address these questions, data from five states was used to model confidence interval and cell-size combinations. The study used a single year of elementary/middle school mathematics and reading achievement test data from five states, modeling selected minimum cell sizes from 10 to 100, and confidence interval sizes from 70% to 99%.

Increases in minimum cell sizes for the special education subgroup were associated with a large increase in the number of schools meeting AYP targets for each of the five states assessed. Increased confidence interval sizes were also associated with an increase in pass rates, but a very much smaller increase. While raising the minimum-n is an effective means of increasing the passing rates of schools, it does so at a considerable cost to special education students in terms of being excluded from the accountability system. When the data were modeled to reflect testing in all grades 3-8, many more special education students’ results are included in the accountability system, assuming that states will not increase the minimum-n. If the implicit theory of action guiding NCLB accountability requirements is to improve instruction and thus outcomes for all students, schools and districts must be accountable for all subgroups in order to ensure that these students are appropriately served. The effect of increasing the minimum-n to exclude substantial portions of special education students must be considered a threat to the validity of the accountability system.

Effect of Minimum Cell Sizes and Confidence Interval Sizes for Special Education Subgroups on School-Level AYP Determinations

Mary Ann Simpson, Brian Gong, and Scott Marion

Center for Assessment

September 22, 2005

Introduction: Judging School Performance under NCLB

The “No Child Left Behind Act” (NCLB) requires that schools be held accountable for the performance of the school as a whole as well as for designated subgroups, beginning with the 2002-03 academic year. Subgroups specified by NCLB include racial/ethnic groups, economically disadvantaged students, students with disabilities, and students with limited English proficiency. States are required to determine whether, for each school, the school as a whole and each subgroup within the school has met a set of Annual Measurable Objective (AMOs) in reading/English language arts and mathematics. In general, the AMOs are the percent of students that must score proficient or above on the state assessments for the school and/or the subgroup to meet AYP. Thus, for example, if the state AMO is 45% in reading, to meet Adequate Yearly Progress, or “AYP”¹ a school would need to have at least 45% of all its eligible students score proficient or above, *and* also have at least 45% of the students in each subgroup score proficient or above: at least 45% of the students with disabilities, 45% of the African-American students, and 45% of its Native American students, and so on. If one group fails to meet the AMO, then the school does not meet AYP.² A school that fails to meet AYP two or more years faces specific sanctions established by NCLB and/or the state. The AMOs under NCLB rise over time until the requirement is 100% of students scoring proficient or above by 2014.

¹ NCLB requires that a judgment be made annually whether every school did or did not “make AYP.” AYP stands for “Adequate Yearly Progress,” which is a term inherited from previous versions of the legislation. In fact, under NCLB schools do not have to make any progress from year to year as long as they are above the AMO.

² Under NCLB, schools have to meet additional requirements in order to meet AYP. For simplicity, in this paper we do not address these other requirements, which include minimum performance on another academic indicator other than test scores—such as graduation rate for high schools; and the requirement of 95% participation on the state assessments.

NCLB Provisions to Support Making Valid and Reliable School Decisions

The NCLB statute and regulations stipulate that states must make reliable and valid decisions regarding whether or not schools have met AYP. The law provides some provisions intended to support making reliable and valid decisions. For example, a school must fail to meet AYP for two years in a row before it is subject to some sanctions; this provision is a partial safeguard against the unreliability caused by any “good class, bad class” fluctuations in the sample of students from one year to the next.

While NCLB specifies that a school must fail to meet AYP two years in a row, NCLB regulations give state the flexibility to make a number of additional decisions that affect the reliability and validity of the state’s version of the accountability system, subject to review and approval by the United States Department of Education (USED). Most states have focused on improving the decision consistency, i.e., the reliability, of the identification decisions. Two common approaches states have had approved to address concerns about reliability are to use a “minimum cell size” and to use confidence intervals (Marion, White, Carlson, Erpenbach, Rabinowitz, and Sheinker, 2002). Every state has set minimum cell sizes, and approximately 40 states are using confidence intervals. Across the nation, states have set minimum cell sizes that range between 10 and 80 students or more (Forte Fast & Erpenbach, 2004).³ According to NCLB rules, if a school does not have the minimum number of students for a subgroup calculation, that subgroup is treated as “meeting AYP” for the purposes of determining whether the school met AYP.

In addition to setting a minimum cell size to insure statistical reliability by accounting for year-to-year fluctuations due to sampling error, states may employ a confidence interval to say that a school’s observed performance was truly below the AMO with a specified degree of confidence. USED has approved proposals from a majority of states for either a 95% or 99% confidence interval (Forte Fast & Erpenbach, 2004), meaning that they are willing to accept errors 5% or 1% of the time in stating that a particular subgroup in a school⁴ did not meet AYP when it truly did. In practice, states have implemented a one-sided confidence interval that

³ Some states use a percentage, such as 15 % of the enrolled students. In a large high school, this could be the equivalent of a hundred students or more.

⁴ Since AYP is determined for most schools as a result of multiple decisions, the actual error rate can be considerably more than the nominal (e.g., 5% or 1%) error rate.

focuses on avoiding identifying schools as not having met AYP if they truly have. If a school's or subgroup's observed performance (e.g., percent proficient) falls within the confidence interval or higher, then the school/subgroup is counted as meeting the AMO.

On the other hand, states have not attended to the validity requirements to the same extent as they have for reliability issues (Marion & Gong, 2003). Separating reliability and validity, as many measurement professionals have been telling us for a long time, is a false distinction. Many of the so-called reliability solutions such as raising the minimum-n have considerable validity implications. In general, accountability system validity focuses on the accuracy of the identification of schools (i.e., are the "right" schools being labeled as passing or failing?), the consequences—both positive and unintended negative—of the accountability system, and the subsequent interventions as a result of identifying schools (Marion & Gong, 2003). One of these validity implications is central to this paper: the consequences for special education students as a result of being included or excluded in the accountability system.

Focus on Students With Disabilities

Special education students are an important subgroup educationally and for school assessment and accountability systems. This was true prior to NCLB, especially with the advent of IDEA 1997, and the NCLB law mentions students with disabilities specifically as one of the subgroups for which schools are to be held accountable. NCLB has caused intense discussion around issues of how appropriately to assess students with disabilities and include them in the accountability system. Students with disabilities have become very practically and politically significant in the early years of NCLB implementation. Many states are finding that a high proportion of schools are not meeting AYP because students with disabilities tend to contribute to schools' failure to meet AYP at a substantial rate (Bielinski & Ysseldyke, 2000; Thurlow, Nelson, Teelucksingh, & Ysseldyke, 2000; Thurlow & Wiley, 2004). One view is that this finding is accurate and valid—in fact, the performance of students with disabilities is substantially lower than other subgroups'. Nevertheless, many state leaders have, for a variety of legitimate reasons, expressed concern about the potentially high number of schools identified as not meeting AYP. Among other strategies, this has resulted in states searching for ways to decrease the potential impact of the students with disabilities subgroup on AYP determinations.

One method being employed to reduce the impact of subgroups on school identification has been increasing the minimum cell size, either in general or for the special education subgroup specifically. A growing number of states are also using confidence intervals and seeking to increase the width of the confidence bands to improve the reliability of the AYP decisions (e.g., from 95% to 99%). Although states' concern with potential over-identification of schools is understandable, if a substantial number of schools "meet AYP," but do so without actually including their special education subgroup in the calculations, the intention of the law is being circumvented, and students may not be receiving needed attention.

Focus of Study and Analysis Methods

This study addresses three questions:

- First, considering the full group of students and the special education subgroup, what is the likely effect of minimum cell size and confidence interval size on school-level AYP determinations? That is, as minimum cell size and confidence interval size vary, how much change takes place in percentage of schools identified as not meeting AYP? The study examined selected minimum cell sizes from 10 to 100, and confidence interval sizes from 70% to 99%.
- Second, what effects do the changing minimum cell sizes have on inclusion of special education students, especially for schools that are declared as "meeting AYP"? As minimum cell sizes increase, more schools will not have enough special education students to meet the minimum cell size. How large is this impact on schools and on the special education population in the state? The effect of confidence intervals vary by group size (e.g., all things being equal, the confidence intervals are wider for smaller groups than larger groups), but confidence intervals do not eliminate any size group from consideration. Therefore, these analyses did not apply to varying confidence interval sizes.
- Third, with the NCLB requirement that schools assess grade levels 3-8 in their AYP calculations beginning in the 2005-2006 academic year⁵, what is the likely effect of including these additional students in school-level AYP determinations?

⁵ Most states assessed one grade per grade span prior to NCLB, i.e., once in elementary, middle, and high school. NCLB requires that states assess annually in grades 3-8, and once in grades 9-12, for math and English language arts/reading starting 2005-06.

To address these questions, varying combinations of confidence interval and cell-size were analyzed using actual achievement data from a small, but varied set of states. The study used a single year of elementary/middle school mathematics and reading achievement test data from five states.⁶

Student-level achievement data for reading and mathematics were analyzed for each state. Each student was declared proficient or not proficient in reading and mathematics according to that state's rules⁷. The percent of students proficient was calculated for each school in math and reading, for both all the students (assessed) in the school (referred to as "the school as a whole") and for the special education students (assessed) in the school. A school was deemed meeting AYP if the percent proficient for both reading and mathematics exceeded a given state's AMOs for both reading and mathematics for the school-as-a-whole and for the special education students *or* if the percent proficient in both reading and mathematics exceeded the state's AMOs for reading and mathematics for the entire participant pool, and the special education subgroup did *not* meet the minimum cell size for inclusion in the calculations.⁸ Passing rates were calculated for minimum cell sizes of 10, 20, 30, 60, 80, and 100 students. Additionally, passing rates were calculated for each of these cell sizes when 75, 90, 95, and 99 percent confidence intervals were implemented.

Basic information about schools and students in the five states' data sets is shown in Table 1. Of the five states, three are small and the other two are moderate size (approximately 50,000 students tested per grade level). Two states included every grade level in their accountability tests (states 4 and 5). The proportion of tested participants in grades 3-8 who were special education students ranged from a low of approximately 11 percent to a high of approximately 20 percent. This range bracketed the national average of approximately 12% special education students. The average number of students per school in grades 3-8 ranged from fewer than 20 to more than 300.

⁶ Either 2003 or 2004 data were analyzed, depending on availability and other factors, such as the stability of the state's accountability policies.

⁷ Appendix 1 gives details of each state's proficiency levels and mathematics and reading achievement scales.

⁸ This study did not try to replicate the states' actually final AYP results, which would involve complex inclusion rules, consideration of academic indicators other than test scores, participation rates, and other elements, especially appeals, required by NCLB and that vary across the states.

Table 1

Basic Information on States Included in Analysis

State	Region	Year	Number of tested students in grades 3-8	Percent of tested students in special education	Average number of tested students per school (standard deviation shown in parentheses)	Grade levels included in accountability calculations (elementary and/or middle schools)
1	Northeast	2003	25,857	20.0%	92.4 (18.2)	4, 8
2	Southeast	2003	114,165	14.6%	88.8 (12.9)	4, 8
3	Northwest	2004	129,471	11.5%	117.1 (84.9)	3, 5, 8
4	Northwest	2003	61,816	13.7%	18.9 (24.2)	3-8
5	West	2003	222,484	11.0%	307.7 (237.0)	3-8

The AMOs for the five states represented a large range (see Table 2)—36 percentage points between the lowest and highest AMOs in reading and 32 percentage points in math. The lowest AMO in reading was 40%, while the highest was 76%. In general, the math AMOs were lower than reading, but exhibited a similar range of differences across the five states, with the lowest math AMO equal to 30% and the highest equal to 62%. The states ranked the same for reading and math AMOs (i.e., a state with a relatively lower AMO in reading had a relatively lower AMO in math), with one exception: State #1’s middle school math AMO was lower compared to the other states relative to its ranking based on reading AMOs. The AMOs were determined by each state according to the percent of students proficient in the school containing that state’s “20th percentile student,” following a specific methodology mandated by NCLB (PL 107-110, Section 1111). One state (state number 1) used index scores ranging from 0-100 to express school performance, rather than a percent proficient. This state’s AMOs were also expressed on this scale.

Table 2

*Annual Measurable Objectives (AMOs) for Elementary and Middle Schools in Each State
(Percent proficient unless otherwise stated)*

State	Year	Reading	Mathematics
1	2003	76.1 (elementary schools)*	61.7 (elementary schools)*
		68.0 (middle schools)*	46.1 (middle schools)*
2	2003	40%	30%
3	2004	40%	39%
4	2003	64%	55%
5	2003	65%	57%

* State number 1 employed school performance scores on a 0-100 metric for each school. Additionally, the state created separate AMOs for elementary and middle schools.

Table 3 shows the percent of students proficient in ELA and Mathematics by special education status for each of the five states.

Table 3

Percent of Students Proficient or Mean School Performance Score in Reading and Math for School-as-a-Whole and for Special Education Subgroup

State	Year	School-As-A-Whole		Special Education	
		ELA	Mathematics	ELA	Mathematics
1	2003	88.59(14.67) ^e	81.12 (17.94) ^e	76.58(17.73) ^e	71.53(19.74) ^e
		84.38(14.80) ^m	70.39(23.95) ^m	70.33(16.03) ^m	52.62(21.12) ^m
2	2003	58.6 %	57.1 %	25.5 %	28.6 %
3	2004	71.0 %	69.8 %	33.3 %	37.1 %
4	2003	70.4 %	65.4 %	34.6 %	30.0 %
5	2003	76.8 %	71.4 %	33.7 %	33.8 %

^e Mean school performance “index score” for elementary schools

^m Mean school performance “index score” for middle schools

Results

Analyses of Actual Data

School Identification Rates as a Result of the Special Education Subgroup

The first set of analyses examined the simple descriptive statistics comparing the percentage of schools that meet the AMOs for the school as a whole subgroup⁹ and for the special education subgroup (see Table 4). Notably, the pass rate for schools with regard to special education is quite low compared to the school as a whole. In other words, the performance of the special education subgroup will lead to schools’ failure at a noticeably higher rate than for the school as a whole. The final column of table 4 shows the percentage of schools reaching AMOs for the student body as a whole, but lacking sufficient cell sizes to assess the progress of special education students. Several details of this table bear mentioning. In the five states studied, over 80 percent of schools that passed their subgroup AMO did so without

⁹ We acknowledge that it seems ironic to call the “school as a whole” a subgroup, but that is a specific NCLB defined-subgroup.

assessing the proficiency of their special education students. An additional finding from these analyses is the variability in passing rates (minimum approximately 46 percent, maximum approximately 92 percent). The two states with the lowest passing rates (states 4 and 5) are the two states currently testing every grade. Again, these results are averaged across all minimum cell sizes and confidence intervals.

Table 4

Percent of Schools Meeting AMOs for Particular Student Subgroups Across All Experimental Conditions.

State	Passed: School-As-A-Whole (% of schools)	Passed: Special Education (% of schools)	Passed * (% of schools)	% of Total Schools that Passed but Lacked the Minimum <i>n</i> in Special Education
1- (<i>n</i> = 277)	95.5%	67.4%	89.6%	84.1%
2- (<i>n</i> = 1283)	85.4%	36.9 %	77.9 %	92.7 %
3- (<i>n</i> = 1112)	96.0 %	49.3 %	87.9 %	90.4 %
			88.5 %	89.9 %
4- (<i>n</i> = 440)	61.6%	13.3%	46.3%	93.7%
5- (<i>n</i> = 723)	78.8 %	9.8 %	50.8 %	92.3 %

*Passed Both Components *or* Passed School-As-A-Whole But Lacked Minimum *n* in Special Education

The Effect of Minimum-*n*

The number of students required to define a set of students as a group has been one of the most discussed aspects of states' implementation of AYP calculations. It has been argued previously (e.g., Marion, et al., 2002) that minimum-*n* is much less of a reliability issue than a consequential validity concern. The analyses presented in Table 5 document the effects, while holding all other aspects of states' accountability plans constant, of altering the minimum number of students necessary to constitute a subgroup on the percent of schools passing AMOs for each of the five states. As one would expect, an increase in the minimum cell size was

associated with an increase in the percentage of schools passing AMOs. All but one state (state 1) showed a difference of more than 25 percentage points. This is likely due to the relatively high pass rates in this state..

Table 5

Percent of Schools Meeting AMOs by Minimum Cell Size

State	Minimum Cell size					
	10	20	30	60	80	100
1	77.5%	86.2%	90.1%	93.7%	95.1%	95.1%
2	58.0%	73.0%	80.2%	85.1%	85.2%	85.4%
3	70.8%	82.1%	90.4%	95.8%	96.0%	96.0%
4	28.2%	35.0%	40.9%	56.3%	57.7%	59.5%
5	18.4%	26.2%	39.8%	70.6%	74.0%	75.8%

Consequences of Increasing Minimum-n

Two analyses were conducted to examine the consequences on special education students of increasing the minimum-n. The first demonstrates quite conclusively for these states that as the cell size requirements increase, fewer schools are held accountable for ensuring that their special education students meet the AMOs. Table 6 shows, for each minimum cell size, the percentage of schools passing their AMOs but without sufficient numbers of special education students to assess their performance. When minimum cell sizes approached 60, almost 100 percent of schools in all five states were able to “pass” AYP without the performance of special education students as an identifiable subgroup taken into account.

The second analysis focuses on the percentage of special education students that would be excluded¹⁰ from the accountability system as a function of increasing cell size. Table 7 shows the percentage of tested special education students excluded from the AYP calculations for each state and cell size. For the three states not currently testing in every grade, more than one-third of special education students were excluded from AYP calculations at a minimum cell size of 20.

¹⁰ We recognize that these students are not fully excluded because they count in the whole school calculations, but practically for most AMO levels, schools could feasibly ignore the performance of special education students until 2011 or so.

For these states, by the point the minimum cell size reached 60 students, nearly 100 percent of special education students were not included in the AYP calculations. This has consequences for special education students and for the validity of the accountability system.

Table 6

Percent of Passing Schools Not Having Enough Special Education Students to Meet Minimum Cell Size Requirements

State	Minimum Cell size					
	10	20	30	60	80	100
1-	36.7%	76.8%	84.1%	97.7%	1.0%	1.0%
2-	57.8%	90.8%	96.6%	100.0%	100.0%	100.0%
3-	51.6%	81.2%	95.7%	100.0%	100.0%	100.0%
4-	70.8%	83.6%	91.6%	99.7%	100.0%	100.0%
5-	45.8%	69.6%	89.2%	99.3%	99.8%	99.9%

Table 7

Percent of Special Education Testing Participants Excluded By Minimum Cell Size

State	Minimum Cell size					
	10	20	30	60	80	100
1-	10.3%	38.5%	49.6%	86.2%	97.7%	97.7%
2-	18.5%	54.1%	75.7%	98.6%	98.9%	100.0%
3-	10.7%	41.2%	73.7%	99.1%	100.0%	100.0%
4-	8.7%	20.7%	31.6%	72.4%	79.7%	87.0%
5-	1.5%	6.9%	20.3%	67.5%	79.9%	87.5%

The Effect of Confidence Intervals on AYP Pass Rates

One approach that has been advocated for improving the reliability of AYP decisions has been to use confidence intervals around either the AMO or the school's observed score (e.g., Hill & DePascale, 2003; Marion, et al., 2002). In these analyses, the effect of confidence interval size is reported at the average of the minimum cell sizes (Table 8). The passing rate increases with increasing confidence interval size, however, the increase is smaller than that observed for minimum cell sizes (see Table 8)¹¹.

Table 8

Percent of Schools Passing AMOs by Confidence Interval Size

State	Confidence Interval Size				
	NONE	75	90	95	99
1-	86.1%	88.4%	90.5%	90.7%	92.4%
2-	68.1%	73.6%	79.3%	81.3%	86.9%
3-	83.1%	86.3%	89.0%	90.8%	93.3%
4-	37.8%	42.8%	46.7%	49.2%	55.1%
5-	45.8%	48.3%	51.4%	52.6%	55.9%

Projections for Testing Every Grade 3-8

States are required to test every grade, 3-8 and once in high school, by the 2005-2006 school year. Prior to that year, schools were required to test students once each in elementary, middle, and high school. With fewer grades being tested, there are fewer students eligible to meet minimum cell sizes. Further, confidence intervals vary inversely as a function of sample size (i.e., they are wider when sample sizes are smaller). Therefore, if the level of the confidence interval does not change, they will, by definition, be narrower when more students are included in the system. Similarly, with more grades tested, more subgroups will meet the minimum-n threshold (assuming it is maintained at the same level). The analyses presented in this section

¹¹ Appendix 2 describes the inferential statistical analyses underlying conclusions presented in this report.

project how the various design decisions play out when the full assessment system is implemented.

Three of the five states (states 1, 2 and 3) did not test every grade in recent years. Data from these states' October, 2004 enumeration of their schools' enrollments was used to make projections of passing rates likely when every grade, 3-8, is tested. It was assumed that the untested students were sampled from the sample population as tested students and, therefore, the percent proficient for the tested and untested groups was identical. It was also assumed that the proportion of special education students was the same between the tested and untested grades. Each school's total enrollment, grades 3 -8, was used as the participant count for analyses by minimum cell size and as sample size in the calculation of the confidence intervals for the analyses by confidence interval size.

Tables 9-10 show the estimated numbers of students and passing rates for the three states currently testing two or three grades when they were projected to test every grade in grades three through eight. Table 11 shows the differences in pass rates from partial to every grade testing for these three states. As one would expect, the pass rates for the student body as a whole did not change very much from partial to complete grade testing. However, the overall pass rate decreased between approximately 7-20 percent.

Table 9

Projected Average Number of Testing Participants Per School If Every Grade Tested.

State	Projected Mean (Standard Deviation) Number of Students Participating in Testing	Projected Mean (Standard Deviation) Number of Special Education Students Participating in Testing
1- (<i>n</i> = 244)	294.83 (59.22)	59.22 (48.11)
2- (<i>n</i> = 1230)	267.01 (206.12)	39.02 (30.12)
3- (<i>n</i> = 1012)	248.63 (210.89)	29.06 (24.67)

Table 10

Projected Percent of Schools Passing AMOs for Particular Student Subgroups Across All Experimental Conditions If Every Grade Tested.

State	Passed: School-As-A-Whole (% of schools)	Passed: Special Education (% of schools meeting minimum-n)	Passed * (% of all schools)	% of Total Schools that Passed but Lacked the Minimum <i>n</i> in Special Education
1- (<i>n</i> = 244)	98.0%	67.3%	80.9%	53.9%
2- (<i>n</i> = 1230)	81.9 %	32.8 %	58.2 %	75.5%
3- (<i>n</i> = 1012)	96.3%	36.8%	75.8%	83.5%
4- (<i>n</i> = 440)	61.6%	13.3%	46.3%	93.7%
5- (<i>n</i> = 723)	78.8 %	9.8 %	50.8 %	92.3 %

* Actual data from states four and five repeated for ease of comparison. “Passing” in this column refers to those subgroups actually meeting the AMO or not having enough students to constitute a subgroup.

Table 11

Projected Difference in Percent of Schools Passing AMOs Across All Experimental Conditions.

State	Passed: School-As-A-Whole (% of schools)	Passed: Special Education (% of schools meeting minimum-n)	Passed * (% of schools)	Passed but Lacking Minimum <i>n</i> in Special Education (% of passing schools)
1- (<i>n</i> = 277)	2.50%	-0.10%	-8.70%	-30.20%
2- (<i>n</i> = 1283)	-3.50%	-4.10%	-19.70%	-17.20%
3- (<i>n</i> = 1116)	0.30%	-15.80%	-12.70%	-6.40%

*Passed Both Components *or* Passed School-As-A-Whole But Lacked Minimum *n* in Special Education

Effects of Minimum-n with All Grades Testing

As more students are added into the system, more schools will meet the minimum-n thresholds for various subgroups¹². The pattern of projected percentages of schools passing AYP at varying levels of minimum cell size (see Table 12) is similar to the pattern for testing fewer students (see Table 5), although slightly fewer schools are able to pass with more students included. Even with the additional students included in the system, a majority of the projected passing schools do so without having sufficient numbers of special education to constitute a subgroup once the minimum-n reaches 30 students (see Table 13). Likewise, once the minimum-n reaches 20 or 30 students, significant percentages of special education students are excluded from the accountability system even with all grades tested (see Table 14). Figures 1-3 show the exclusion rates for the three states without a full assessment system now compared with the exclusion rates when the system is fully built out as a function of cell size.

Table 12

Projected Percent of Schools Passing AMOs by Minimum Cell Size If Every Grade Tested

State	Minimum Cell size					
	10	20	30	60	80	100
1-	68.4%	69.8%	75.6%	87.1%	91.8%	92.8%
2-	35.1%	41.9%	51.0%	68.8%	74.4%	77.8%
3-	50.9%	59.6%	73.6%	86.5%	90.4%	93.7%
4*	28.2%	35.0%	40.9%	56.3%	57.7%	59.5%
5*	18.4%	26.2%	39.8%	70.6%	74.0%	75.8%

* Actual data from states four and five repeated for ease of comparison

¹² Some states have considered approaches to calculate AYP for each individual grade, in which case adding more grades will not necessarily affect the current minimum-n. We have never considered this a reasonable approach, because it has always been clear to us that the unit of analysis described in the law is the school, not the grade. For example, units that do not meet AYP for two consecutive years go into *SCHOOL* improvement, not *GRADE* improvement.

Table 13

Projected Percent of Passing Schools Not Meeting Minimum Cell Size Requirements for Special Education Students If Every Grade Tested

State	Minimum Cell size					
	10	20	30	60	80	100
1-	1.2%	8.2%	37.9%	79.4%	83.7%	86.7%
2-	8.1%	35.7%	63.8%	93.8%	96.8%	98.6%
3-	26.0%	56.7%	86.1%	99.0%	99.5%	99.9%
4-*	70.8%	83.6%	91.6%	99.7%	100.0%	100.0%
5-*	45.8%	69.6%	89.2%	99.3%	99.8%	99.9%

* Actual data from states four and five repeated for ease of comparison

Table 14

Projected Percent of Special Education Students Excluded By Minimum Cell Size If Every Grade Tested

State	Minimum Cell size					
	10	20	30	60	80	100
1-	< 1%	1.4%	11.3%	40.1%	49.7%	55.4%
2-	1%	6.4%	18.9%	55.7%	70.6%	81.8%
3-	2.7%	13.0%	37.5%	67.5%	77.8%	88.7%
4- *	8.7%	20.7%	31.6%	72.4%	79.7%	87.0%
5- *	1.5%	6.9%	20.3%	67.5%	79.9%	87.5%

* Actual data from states four and five repeated for ease of comparison

Figure 1

State Number 1: Percent Special Education Students Excluded: Partial Grade Testing Versus Projected All Grades Testing.

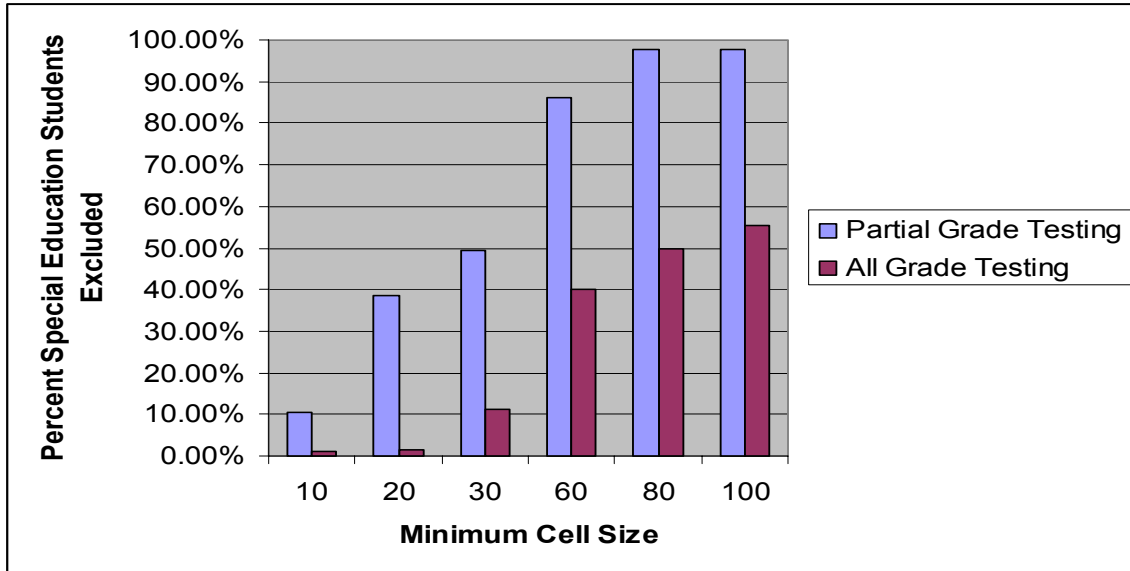


Figure 2

State Number 2: Percent Special Education Students Excluded: Partial Grade Testing Versus Projected All Grades Testing.

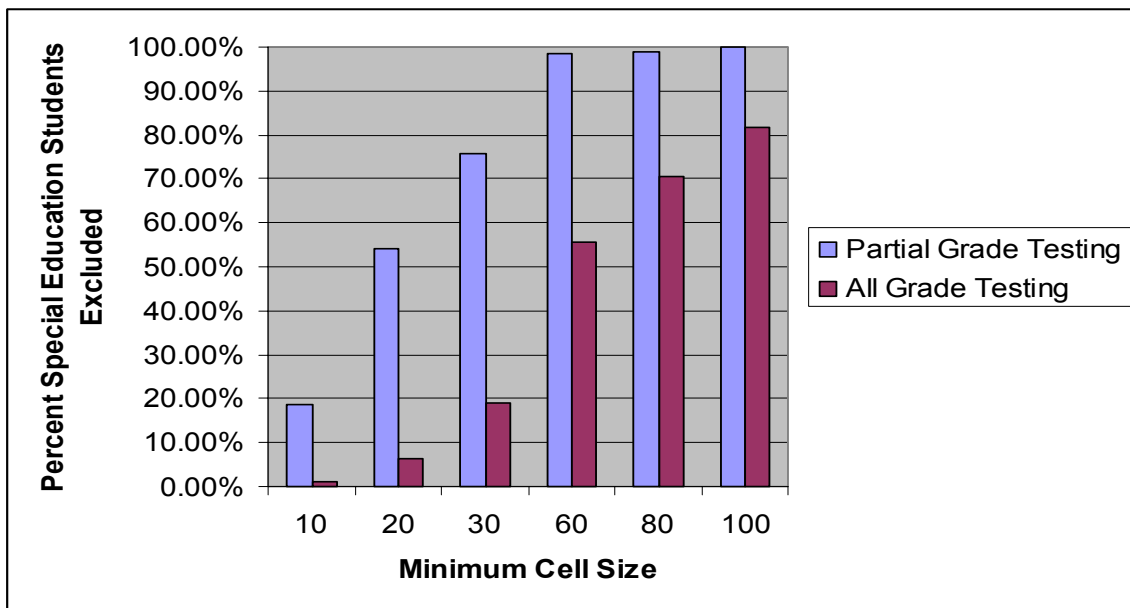
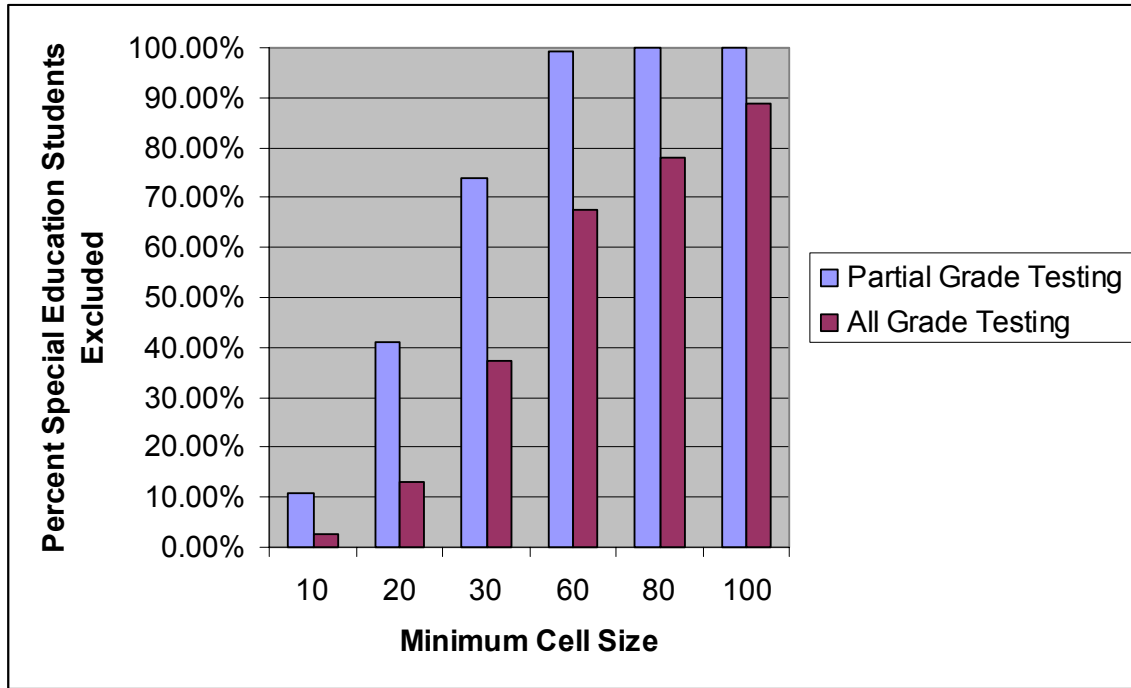


Figure 3

State Number 3: Percent Special Education Students Excluded: Partial Grade Testing Versus Projected All Grades Testing.



Effects of Confidence Intervals with All Grades Testing

When more students are added into the system, the width of the confidence interval bands will decrease. The general pattern found for all grades testing were similar to those from the analyses for partial grade testing (see Table 15).

Table 15

Percent of Schools Passing AMOs by Confidence Interval Size If Every Grade Tested

State	Confidence Interval Size				
	NONE	75	90	95	99
1-	75.8%	78.3%	81.8%	83.4%	85.3%
2-	50.0%	54.5%	58.4%	61.2%	66.5%
3-	70.5%	73.2%	76.3%	77.9%	81.0%

Do the effects of cell size and confidence interval size depend on each other?

Some might be surprised that the main effect of the size of the confidence interval was not larger than it was. This is probably due to the large increase in “free passes” seen as the cell size increased. Table 16 shows the effects of cell size and confidence interval size on the overall pass rate averaged over the current grade testing for states 1-3. Table 17 shows this same information averaged over the projected all grade testing for states 1-3 and current grade testing for states 4-5. From these two tables, it is clear that confidence interval size has its strongest effect when the required minimum n is at its smallest. Furthermore, the effect of required minimum n becomes greatly diminished by the time the confidence interval is increased to 99 percent. Appendix 3 shows the disaggregated results for all states under all conditions.

Table 16

Percent of Schools Passing AMOs by Minimum Cell Size and Confidence Interval Size-Averaged Across States With Partial Grade Testing (States 1-3).

		Minimum Cell Size					
CI Size		10	20	30	60	80	100
	0	55.6%	73.5%	82.1%	87.3%	88.1%	88.1%
	75	62.4%	77.9%	85.2%	90.0%	90.6%	90.7%
	90	70.0%	81.4%	87.8%	92.5%	93.0%	93.1%
	95	74.2%	82.8%	88.4%	93.1%	93.6%	93.6%
	99	81.9%	86.9%	91.1%	94.8%	95.3%	95.3%

Table 17

Percent of Schools Passing AMOs by Minimum Cell Size and Confidence Interval Size-Averaged Across States With All- Grade Testing (Projections for States 1-3, Current Data for States 4-5).

CI size	Minimum Cell Size					
	10	20	30	60	80	100
0	29.4%	37.4%	50.1%	69.6%	73.5%	75.8%
75	34.6%	42.0%	53.2%	72.3%	76.1%	78.5%
90	40.7%	47.0%	56.4%	74.6%	78.3%	80.5%
95	44.5%	50.1%	58.4%	75.4%	79.3%	81.5%
99	51.8%	56.0%	62.8%	77.5%	81.2%	83.4%

Summary and Conclusions

While states have flexibility in meeting the NCLB reliability expectations, their choices can lead to severe consequences for special education students. Most troublesome is the application of high minimum-n requirements. When the minimum-n was simulated to equal 60 students (well within the range of state values), more than half of the special education students in four of the five states—even when projecting all grades testing—were excluded as an explicit subgroup from the accountability system. Clearly, this violates the notion “leaving no child behind,”

Why would state leaders choose a strategy that excludes more than half of an important subgroup? NCLB has put tremendous pressure of states at a time of increased budget pressures and other capacity challenges. It is understandable that state leaders want to reduce the burden of having large numbers of schools failing AYP. Clearly, increases in minimum cell sizes for the special education subgroup were associated with a large increase in passing rates for each of the five states assessed. This increase was due, in large part, to schools’ being less likely to have to include the results for the special education subgroup as the minimum cell size increased. As many suspected, it is considerably easier for a school to meet its AMO without reporting the proficiency of their special education students. Increased confidence interval sizes were also associated with an increase in pass rates, but a very much smaller increase. While raising the minimum-n is an effective means of increasing the passing rates of schools, it does so at a considerable cost to special education students in terms of being excluded from the accountability system. If the implicit theory of action guiding NCLB accountability requirements is to improve instruction and thus outcomes for all students, schools and districts must be accountable for all subgroups in order to ensure that these students are appropriately served. Therefore, tinkering with the minimum-n to exclude substantial portions of special education students must be considered a threat to the validity of the accountability system.

Many more special education students’ data are reflected in the accountability results when all grades are tested. This assumes that states will not increase the minimum-n as more grades are tested. If they do so, then it will likely be a wash between the increase in available students and the loss of these students through increases in required cell sizes.

Although confidence intervals have been suggested as a means of increasing the reliability of school identifications as well as reducing the number of schools failing to make AYP (hopefully because it will reduce those falsely identified), the data presented in this study suggests that confidence intervals have a much smaller impact on AYP pass rates than minimum-n changes. One of the reasons for this finding is the relatively large difference between the observed performance of the special education subgroup and the performance targets in the five states. Three of the five states had relatively high AMOs (e.g., > 60% proficient). If only a small proportion of special education students are scoring proficient, then the confidence intervals will still not be wide enough to overlap the AMO. In other words, if the difference between the percent of special education students scoring proficient and the AMO is large, confidence intervals will still not “help,” assuming the motive for adjustment is to reduce numbers of school identified as not meeting AYP. In only one of the five states did more than 50 percent of the schools have their special education subgroup meet the state’s AMOs.

Some might read this to suggest that confidence intervals are an ineffective tool for addressing reliability and validity concerns with NCLB. However, we read it quite the opposite. Confidence intervals will not help the special education subgroup pass when they should really not pass (i.e., they are far below the AMO), but can help the state leaders make this decision more reliability. On the other hand, minimum-n approaches, as many have been arguing for several years, do little to improve the reliability of subgroup decisions (at least within the range of minimum-n levels being used by most states), but can have severe negative consequences for subgroups excluded and by extension threaten the validity of the accountability system.

References

- Bielinski, J., & Ysseldyke, J. (2000). *Interpreting trends in the performance of special education students* (Technical Report 27). Minneapolis, MN: University of Minnesota, NCEO
- Forte Fast, E. & Erpenbach, W. J. (2004). Revisiting Statewide Educational Accountability Under NCLB: A Summary of Requests in 2003-2004 for Amendments to State Accountability Plans. Washington, D.C.: Council of Chief State Schools Officers.
- 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: A Paper in the Series: Implementing The State Accountability System Requirements Under The No Child Left Behind Act Of 2001. Washington, D.C.: Council of Chief State Schools Officers.
- Marion, S. F. & Gong, B. (2003). Evaluating the Validity of States' Accountability Systems. Paper presented at the Reidy Interactive Lecture Series. October 9-10. Nashua, NH.
- Thurlow, M. L., & Wiley, H. I. (2004). *Almost there in public reporting of assessment results for students with disabilities* (Technical Report 39). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved April 26, 2005, from the World Wide Web: <http://education.umn.edu/NCEO/OnlinePubs/Technical39.htm>
- Thurlow, M., Ysseldyke, J., Nelson, J. R., & Teelucksingh, E. (2000). *Where's Waldo? A third search for students with disabilities in state accountability reports* (Technical Report 25). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- P. L. 107-110 "No Child Left Behind Act of 2001," Title I-Improving the Academic Achievement of the Disadvantaged, Section 1111.

Appendix 1

Details of Each State's Proficiency Scoring for Mathematics and Reading

State	Number of Points on Proficiency Scale	Mastery Determination	Notes
1	Six points [0,1,2,3,4,5] converted to five index levels [0,25,50,75,100].	A student's score is equal to 5.	A school meets AYP if its index score is greater than AMO index score.
2	Five points [1,2,3,4,5]	A student's proficiency score is greater than or equal to 3.	
3	Five points [1,2,3,4,5]	A student's proficiency score is greater than or equal to 4.	This state reports scores for basic reading, reading, writing, mathematics skills, concepts and problem solving. For the current study, the basic reading and mathematics skills proficiency scores were used.
4	Four points [1,2,3,4,5]	A student's proficiency score is greater than or equal to 3.	
5	Four points [1,2,3,4,5]	A student's proficiency score is greater than or equal to 3.	The mathematics test for grades 7-8 may cover algebra, geometry or pre-algebra depending on the student's curriculum. In the current study, a student's score was included regardless of curriculum.

Appendix 2

Inferential Statistical Analyses Conducted for this Report

Separate repeated measures logistic regressions were conducted for each of the five state's passage determinations. SAS, version 8.02, *proc GENMOD* was used (SAS Institute, 2001). The independent variables were minimum cell size and confidence interval size. The logistic regression function in these analyses describes the probability of a school failing. Regression coefficients in the current analyses describe the degree of association between increasing values of the predictor variables with the probability of failing. Cell size and confidence interval size were dummy-coded into a set of dichotomous variables comparing the probability of being declared non-proficient in the very highest level of the variable with that in the other levels. For instance in one state's data, the logistic regression coefficient for a minimum cell size of 10 was 1.82 (.18), $Z = 10.36$, $p < .0001$. This coefficient indicates that a school using a minimum cell size of 10 was approximately 6 times more likely to declared failing than a school with a minimum cell size of 100 special education students.

Regression coefficients comparing the lower minimum cell sizes with the highest minimum cell sizes, and those comparing the narrowest confidence intervals sizes with the widest, were always significantly different from 0. When regression coefficients for the combinations of cell size and confidence interval size were significant, it was usually for the combinations of lowest cell sizes and narrowest confidence intervals. Results were similar for the analyses conducted with projected cell sizes, but the effect sizes for cell size comparisons were much larger than those for confidence interval sizes

Appendix 3

Pass Rates by Cell Sizes and Confidence Interval Size for All States

Table 1

Partial Grade Testing, States 1-3

State							
Cell Size							
1	CI	10	20	30	60	80	100
	0	69.2%	82.6%	87.0%	91.3%	93.3%	93.3%
	75	74.3%	85.4%	88.9%	92.9%	94.5%	94.5%
	90	79.4%	87.0%	90.9%	94.5%	95.7%	95.7%
	95	80.6%	87.0%	90.9%	94.5%	95.7%	95.7%
	99	84.2%	89.3%	92.9%	95.3%	96.4%	96.4%
Cell Size							
2	CI	10	20	30	60	80	100
	0	41.2%	63.6%	72.3%	77.1%	77.1%	77.3%
	75	48.7%	69.0%	77.4%	82.1%	82.2%	82.4%
	90	58.5%	74.4%	81.6%	87.1%	87.2%	87.4%
	95	64.7%	75.9%	82.6%	88.0%	88.1%	88.3%
	99	77.2%	82.3%	87.1%	91.6%	91.6%	91.7%
Cell Size							
3	CI	10	20	30	60	80	100
	0	56.3%	74.3%	86.9%	93.6%	93.8%	93.8%
	75	64.2%	79.2%	89.2%	95.0%	95.2%	95.2%
	90	72.1%	82.7%	90.9%	96.0%	96.2%	96.2%
	95	77.3%	85.4%	91.7%	96.7%	96.9%	96.9%
	99	84.3%	88.9%	93.4%	97.7%	97.8%	97.8%

Table 2

All Grade Testing, States 4-5, Projected States 1-3.

State							
1							
	Cell Size						
CI	10	20	30	60	80	100	
0	58.6%	61.1%	71.3%	84.4%	88.9%	90.2%	
75	63.1%	65.2%	73.4%	86.1%	90.6%	91.8%	
90	69.3%	70.9%	76.2%	88.1%	92.6%	93.4%	
95	73.8%	74.2%	77.5%	88.1%	93.0%	93.9%	
99	77.0%	77.5%	79.9%	88.9%	93.9%	94.7%	
2							
	Cell Size						
CI	10	20	30	60	80	100	
0	23.2%	30.9%	42.3%	62.8%	68.6%	72.2%	
75	28.7%	36.3%	46.5%	66.9%	72.5%	76.0%	
90	34.6%	41.7%	51.1%	69.7%	75.3%	78.5%	
95	39.4%	45.9%	54.3%	71.1%	76.7%	80.1%	
99	49.5%	54.6%	60.6%	73.6%	78.9%	82.0%	
3							
	Cell Size						
CI	10	20	30	60	80	100	
0	37.6%	50.5%	69.7%	84.6%	88.7%	92.0%	
75	45.0%	55.1%	71.6%	85.4%	89.5%	92.9%	
90	52.6%	60.4%	73.6%	86.6%	90.6%	93.9%	
95	56.6%	63.3%	74.9%	87.2%	91.0%	94.3%	
99	62.9%	68.5%	77.9%	88.9%	92.4%	95.6%	
4							
	Cell Size						
CI	10	20	30	60	80	100	
0	17.3%	24.5%	32.0%	49.2%	50.8%	52.6%	
75	23.2%	30.4%	36.9%	53.9%	55.4%	57.2%	
90	28.4%	35.3%	41.0%	57.0%	58.2%	60.1%	
95	31.2%	38.4%	44.1%	59.0%	60.3%	62.1%	
99	41.2%	46.4%	50.8%	62.6%	63.9%	65.7%	
5							
	Cell Size						
CI	10	20	30	60	80	100	
0	10.2%	19.9%	35.3%	67.1%	70.4%	72.2%	
75	13.2%	22.8%	37.5%	69.3%	72.6%	74.5%	
90	18.6%	26.6%	40.1%	71.4%	74.8%	76.6%	
95	21.4%	28.6%	41.4%	71.7%	75.3%	77.0%	
99	28.4%	33.2%	44.8%	73.4%	76.9%	78.9%	