Options for NCLB Designs

Richard Hill
Center for Assessment
2003 RILS

Basics

- NCLB requires both the school as a whole and all subgroups within the school to have satisfactory performance
- Subgroups include
  - All major racial/ethnic groups
  - Students with disabilities
  - Limited English proficient students
  - Students receiving free/reduced price lunch

Basics (cont'd)

- A subgroup is satisfactory only if it passes *either* a status test or an improvement test
- Subgroup must be satisfactory in both mathematics and reading (or ELA)

AYP for Subgroups

- Establish baseline (20th percentile school)
- All subgroups must have either:
  - Status score higher than baseline

    OR
  - Sufficient improvement from previous year (10 percent reduction in percent not proficient)

Judging Subgroups

- "statistically valid and reliable"
- Disaggregation not required when "number of students is insufficient to yield statistically reliable information"
- Choices:
  - Minimum N
  - Confidence intervals
    - Choice of degree of confidence

Volatility of Subgroup Scores

- The results in a particular grade each year operate as though a random sample drawn from a larger population
- "Good class, bad class"
- If the long-term percentage passing for a subgroup is, say, 30 percent, some years the observed results will be higher than 30 percent, and other years it will be less

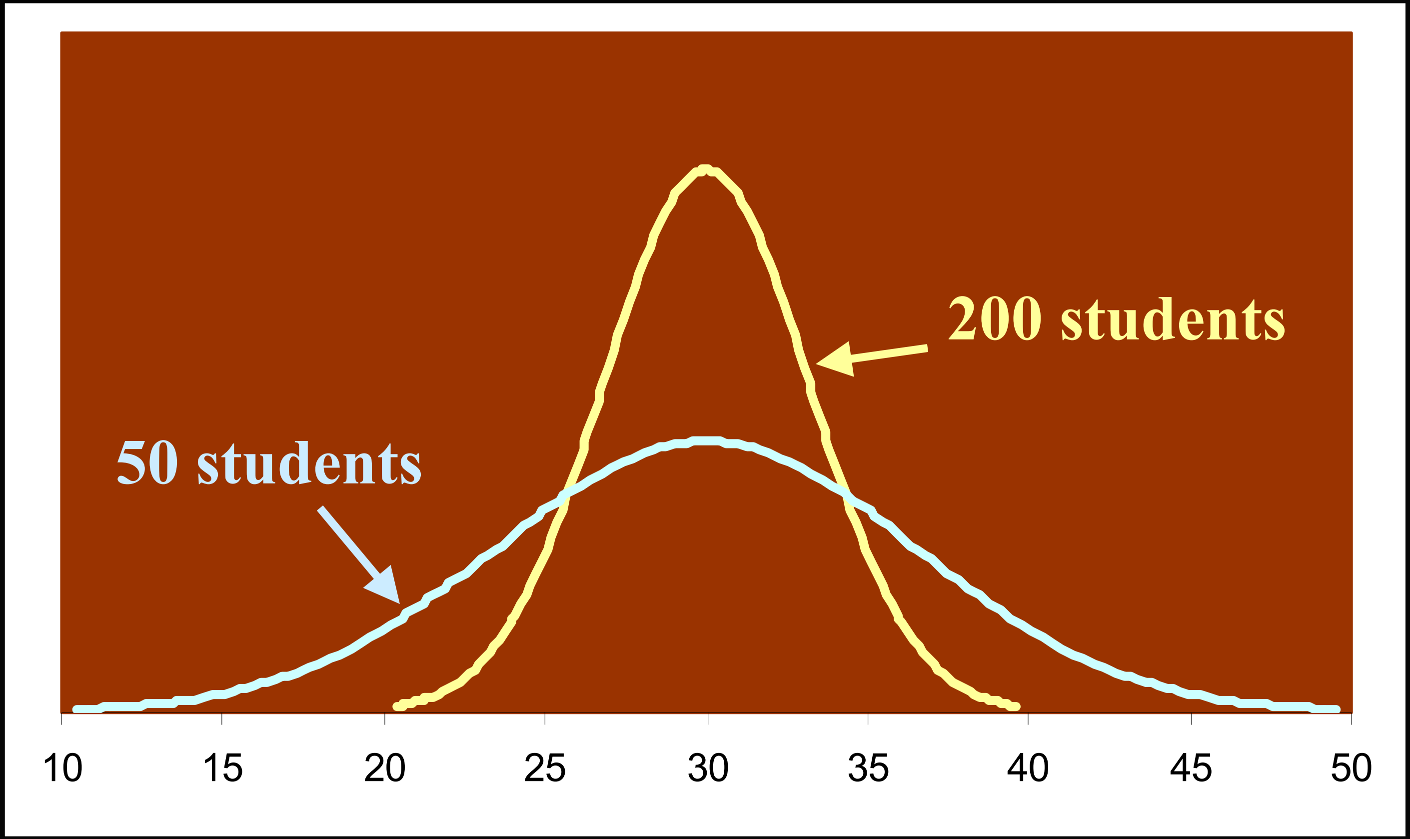
The Basic Issue

- Suppose a subgroup is supposed to have 30 percent of its students passing. After testing, there are three possibilities:
  - 30 percent or more passed
  - Less than 30 percent passed, but close enough to 30 percent that we’re not sure that another sample of students wouldn’t be on the other side of the line (the gray area)
  - So far less than 30 percent passed that we feel confident that the true percentage passing for the school is less than 30

Determining a Reasonable Range of Expected Variation

- If a subgroup's true percentage of students passing mathematics is 30, how often would a random draw from that population produce a sample with only 25 percent passing? 20 percent passing?
- Answer is dependent on the number of students tested

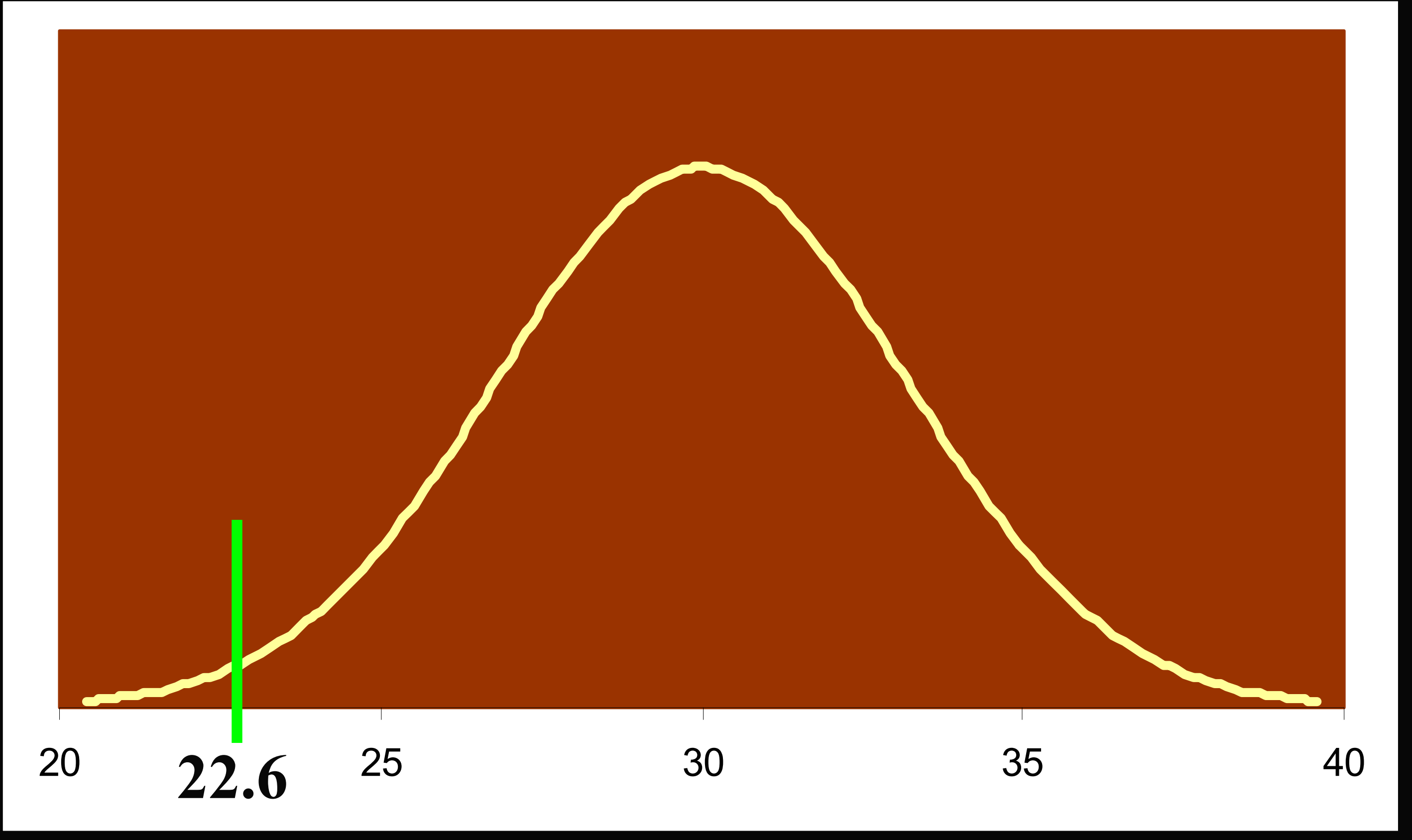
Number Tested Influences Range of Observed Scores



What Score Should We Choose as a Cutoff?

- It should be low enough to convince us that the subgroup is unlikely to truly have a mean at the required level
- The answer will vary depending upon:
  - The degree of confidence we want in the answer
  - The number of students tested
  - The required percent passing
  - The reliability of the test

Choosing a Cutoff When P = 30, N = 200, *alpha* = .01



Hypothesis Testing

- If a subgroup's true mean is 30 and 200 students are tested, it is unlikely (it would only happen 1 time in 100) that it would produce a sample result of less than 22.6
- Therefore, if we observe a sample result less than 22.6 for a subgroup of 200 students, we don't believe it was drawn from a population with a mean of 30 (*we reject the null hypothesis at the .01 level*)

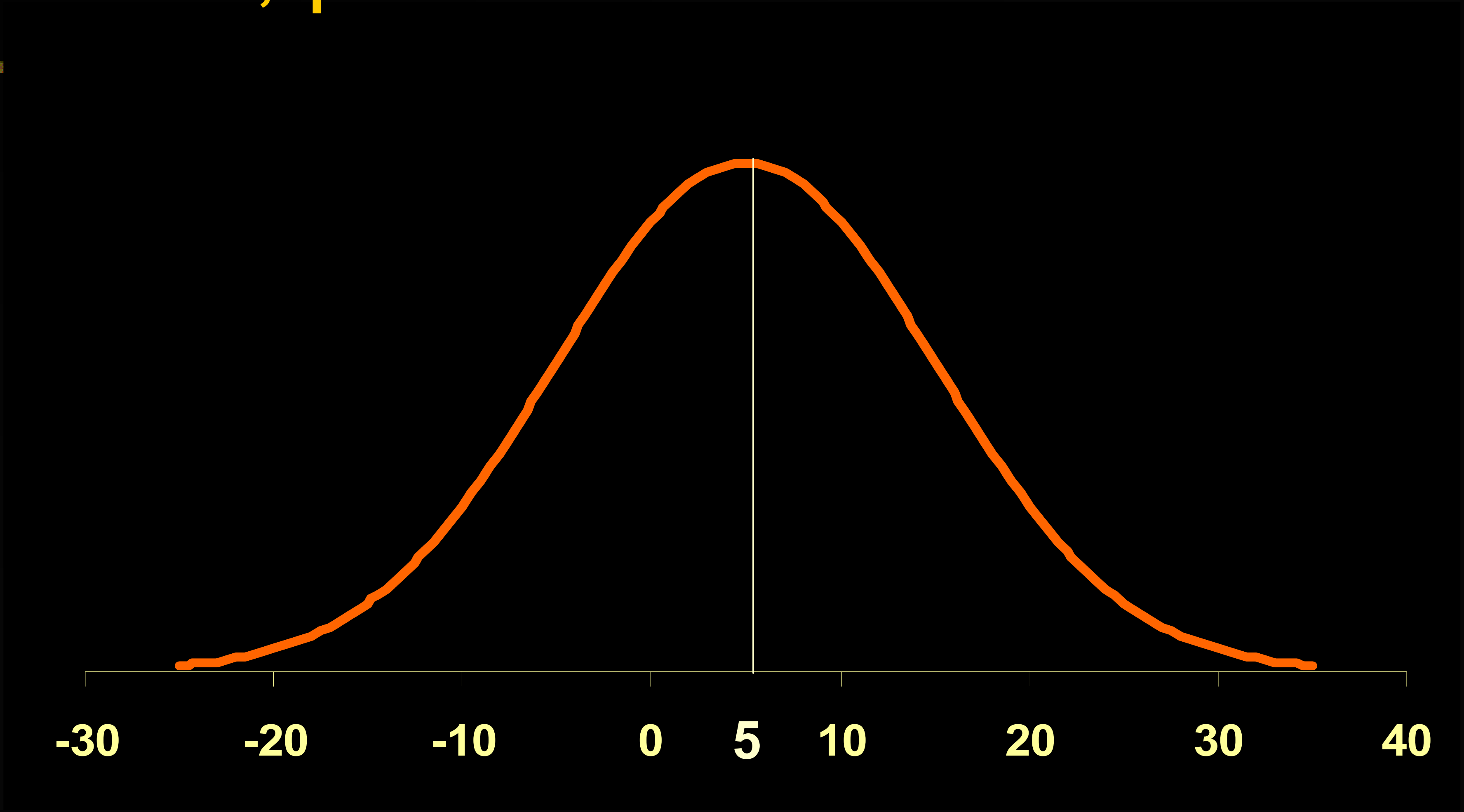
Status vs. Improvement

- Generally can relatively reliably determine status with groups of moderate size
  - One year of error
  - Subgroups often are far from 20th %tile school
- Generally cannot reliably determine improvement even with very large groups
  - Two years of error
  - Amount of improvement expected is relatively small

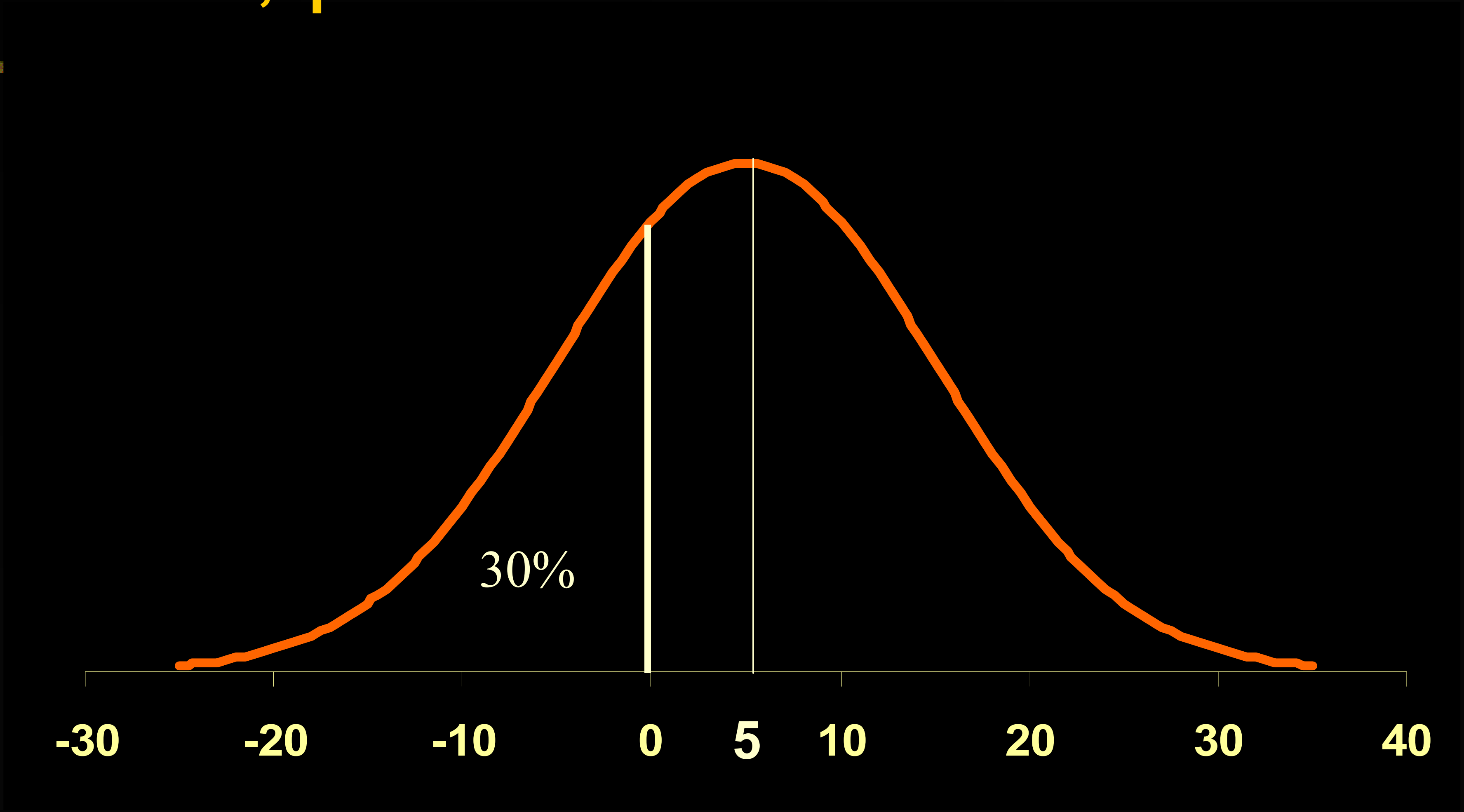
Distribution of Improvement Scores

- If p = .50, groups are required to improve by .05
- If population of school really improves from .50 to .55, what percentage of schools with N = 50 will have observed changes that are 5 percent or more? A *decrease* from previous year?
- What is the bottom 5 percent of that distribution?

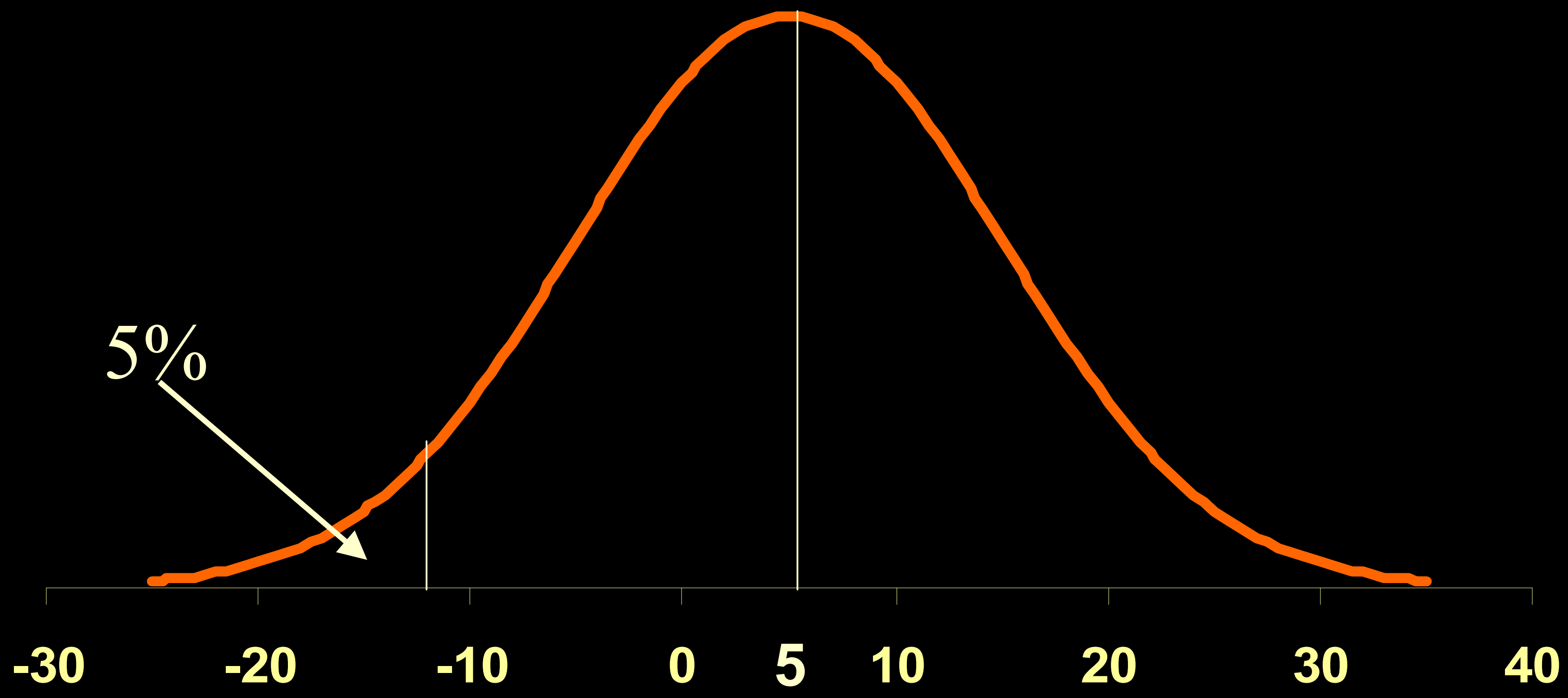
Distribution of Improvement Scores N = 50, p = .55



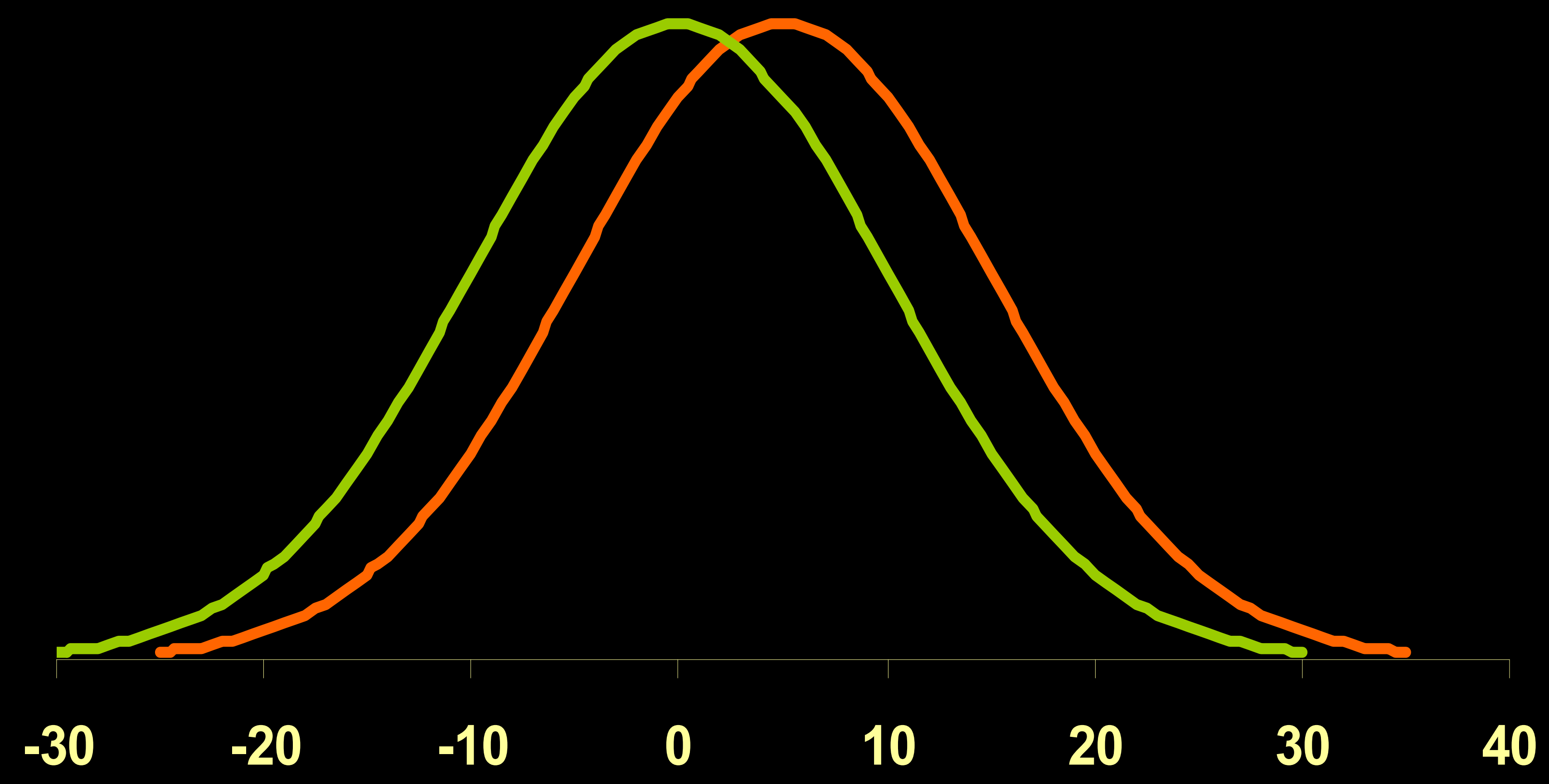
Distribution of Improvement Scores
N = 50, p = .55



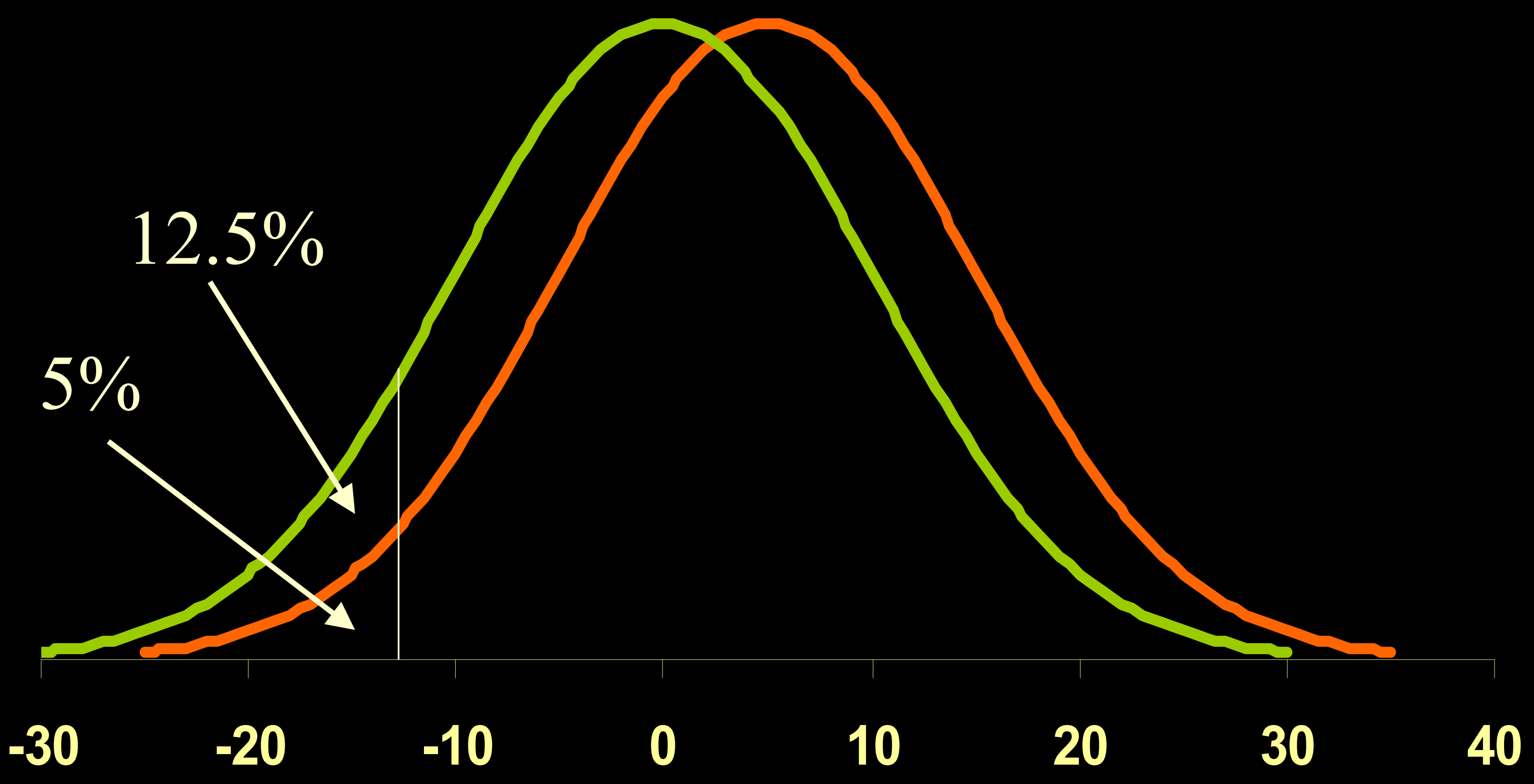
Distribution of Improvement Scores
N = 50, p = .55

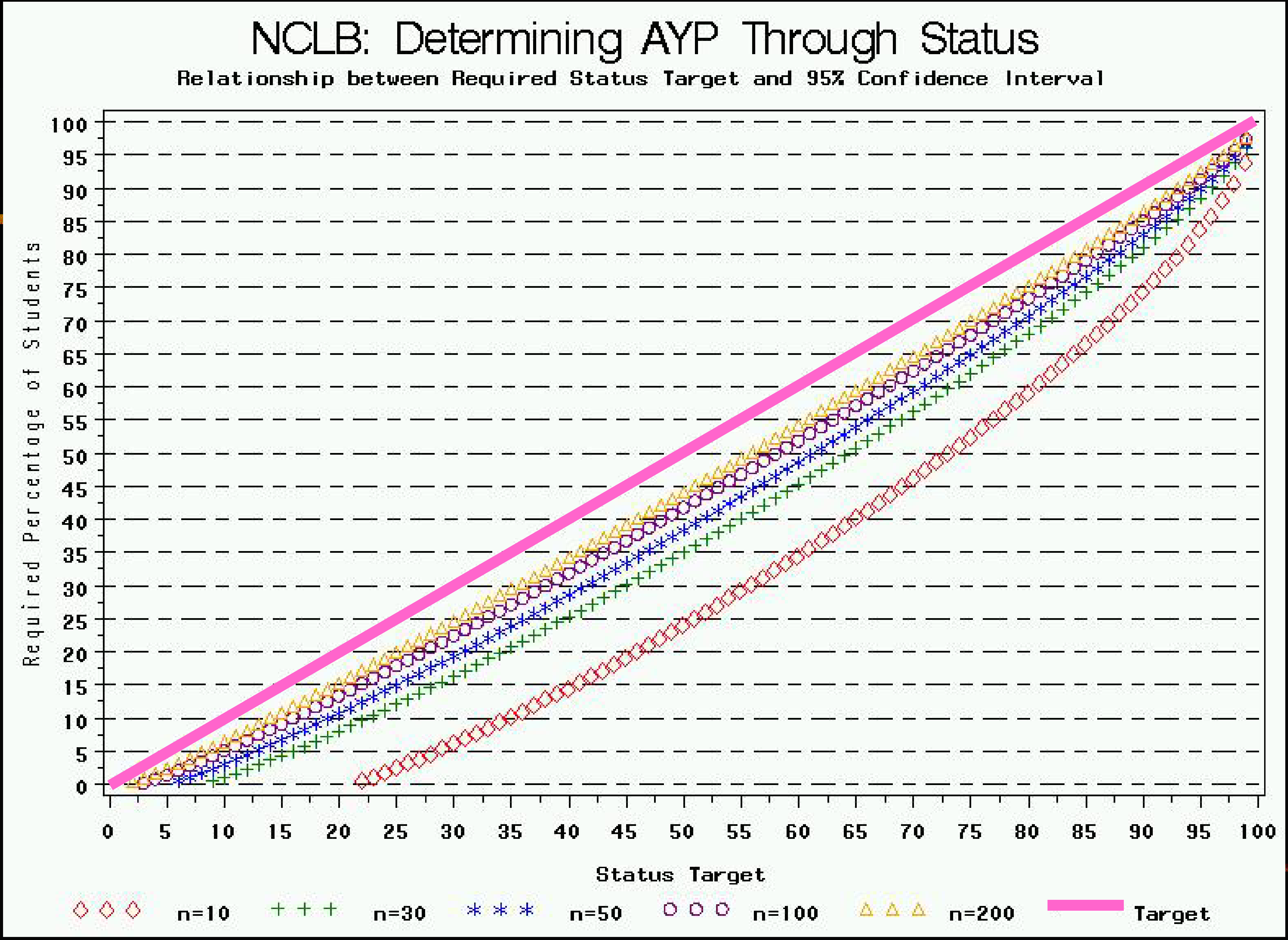


Distribution of Scores
N = 50, p = .50 and .55

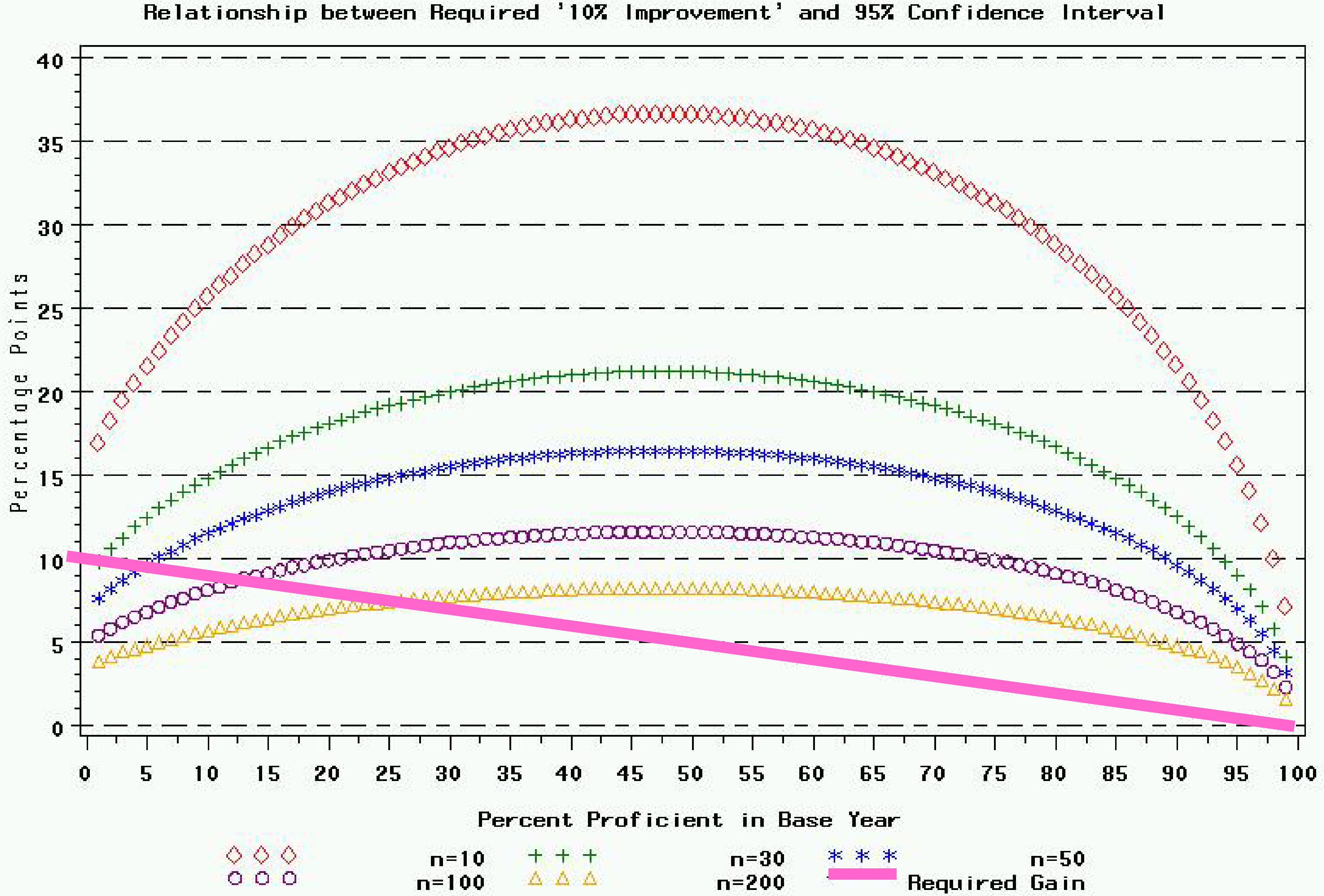


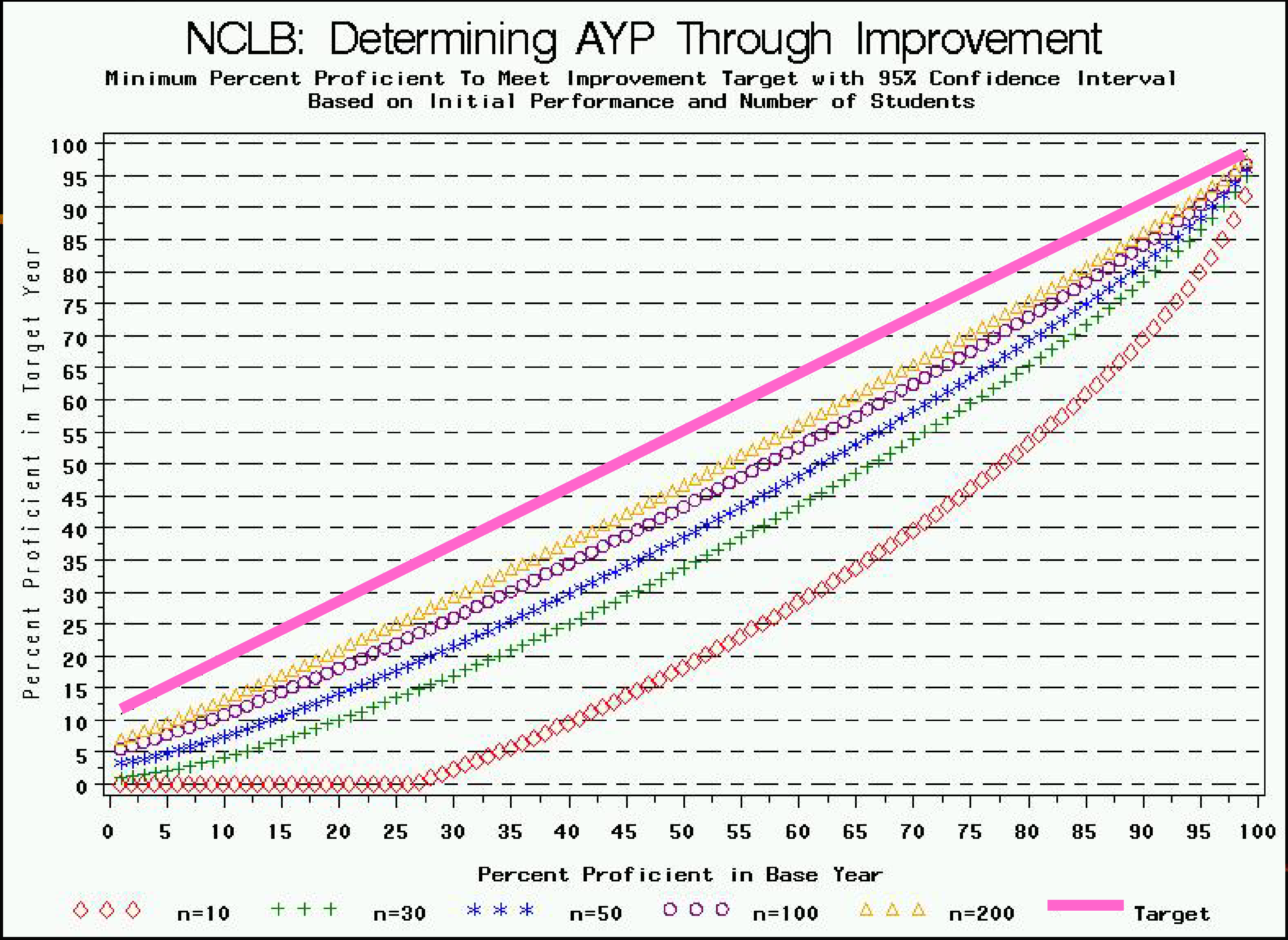
Distribution of Scores
N = 50, p = .50 and .55




NCLB: Determining AYP Through Status
Relationship between Required Status Target and 95% Confidence Interval
Required Percentage of Students
Status Target
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
n=10
n=30
n=50
n=100
n=200
Target

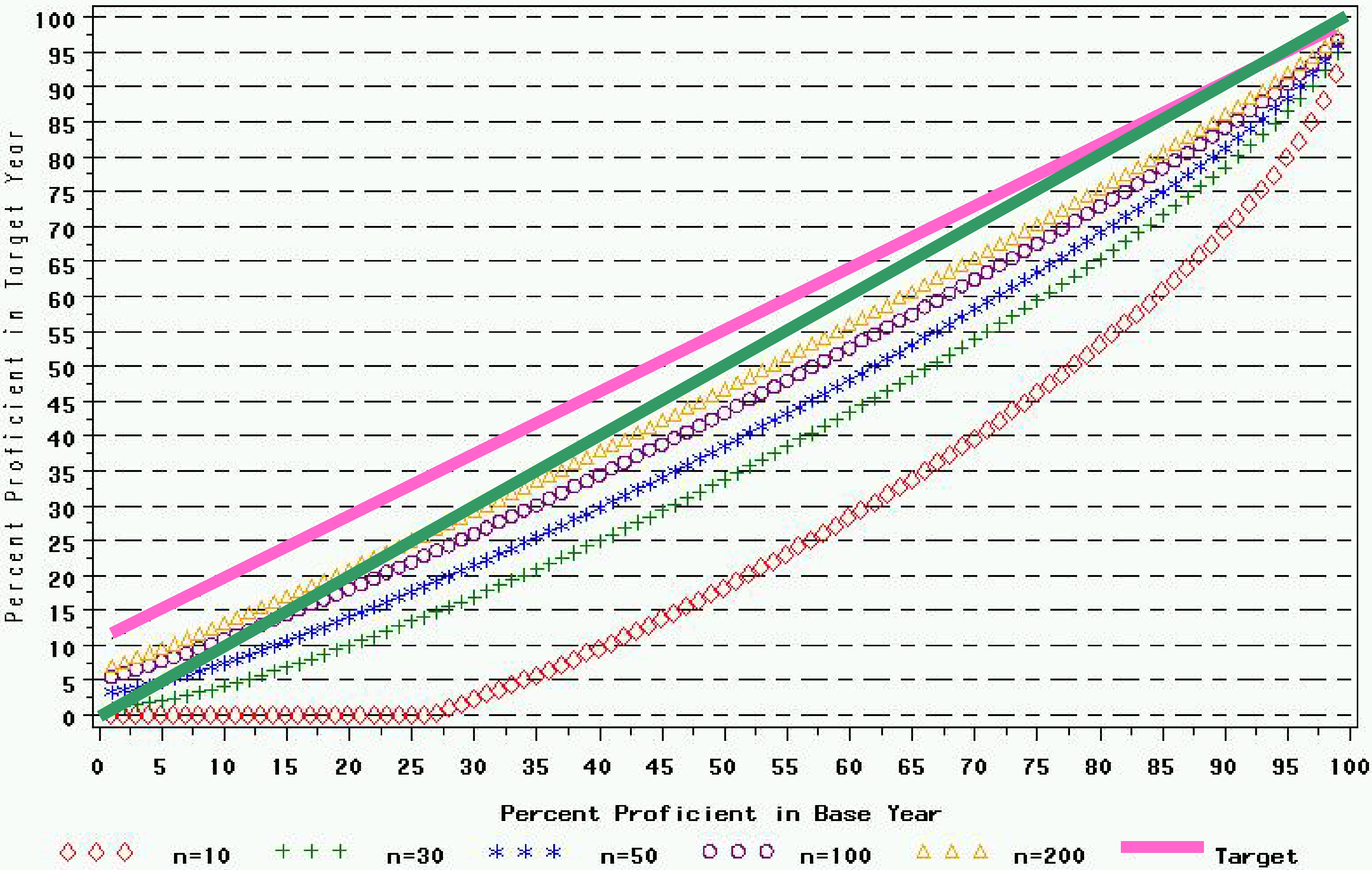
NCLB: Determining AYP Through Improvement




NCLB: Determining AYP Through Improvement
Minimum Percent Proficient To Meet Improvement Target with 95% Confidence Interval
Based on Initial Performance and Number of Students
Percent Proficient in Target Year
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
Percent Proficient in Base Year
n=10
n=30
n=50
n=100
n=200
Target

NCLB: Determining AYP Through Improvement

Minimum Percent Proficient To Meet Improvement Target with 95% Confidence Interval
Based on Initial Performance and Number of Students



Potential Solutions

- Use minimum N
- Set an alpha level higher than .05
- Change decision rules across years
- Take a longer-term look at improvement
- Take a second look at schools within year

Potential Solution 1: Use Minimum N

- Acceptable practice is to set a minimum number (typically 30-50) of students in group
- That practice is both unreliable *and* invalid
  - Unreliable because 30-50 students is an insufficient number to detect improvement
  - Invalid because schools are not held accountable for subgroups with, say, 29 students

Confidence Intervals vs. Minimum N

- Using confidence intervals for improvement means few schools are identified,but those identifications are reliable
- Using minimum N identifies more schools, but just because you've identified *more* doesn't mean you've identified the *right* ones

Potential Solution 2: Higher Alpha

- USED guideline is an *alpha* level of .25
- What *alpha* level should be chosen for each subgroup if the desired *alpha* level for the school is .25 (a *school-wise alpha* level of .25)?
- If 18 tests are run, and all are independent, each test needs to be at the .015 level

Probability of Incorrectly Identifying an Improving School

	Alpha = .05	Alpha = .01
Actual Alpha	.033	.0067

Probability of Incorrectly Identifying an Improving School

	Alpha = .05	Alpha = .01
Actual Alpha	.033	.0067
One Content Area	.19	.05

Probability of Incorrectly Identifying an Improving School

	Alpha = .05	Alpha = .01
Actual Alpha	.033	.0067
One Content Area	.19	.05
Two Content Areas	.32	.08

Potential Solution 3: Change Decision Rules Across Years

- Assumes error in one year is inconsequential—what is important is two consecutive errors
- Choices:
  - Any subgroup, either test
  - Any subgroup, same test
  - Same subgroup, same test

Probability of Incorrectly Identifying an Improving School

	Alpha = .05	Alpha = .01
Any Subgroup, Either Test	.12	.016

Probability of Incorrectly Identifying an Improving School

	Alpha = .05	Alpha = .01
Any Subgroup, Either Test	.12	.016
Any Subgroup, Same Test	.08	.012

Probability of Incorrectly Identifying an Improving School

	Alpha = .05	Alpha = .01
Any Subgroup, Either Test	.12	.016
Any Subgroup, Same Test	.08	.012
Same Subgroup, Same Test	.04	.002

Comparing Improving to Non-improving Schools

	Alpha = .05		Alpha = .01	
	Imp.	Non-imp.	Imp.	Non-imp.
Any Subgroup, Either Test	.12	.38	.016	.13
Any Subgroup, Same Test	.08	.34	.012	.11
Same Subgroup, Same Test	.04	.26	.002	.08

Probability School Labeled As INOI Really Is Improving

	Alpha = .05	Alpha = .01
Any Subgroup, Either Test	.24	.11
Any Subgroup, Same Test	.19	.10
Same Subgroup, Same Test	.13	.02

Potential Solution 4: Look at Improvement over More Years

- Rather than just compare a school's results to last year's, compare to those of two or three years ago

Percentage Passing Subgroups Must Have (1-Year Improvement)

N	0		10		20		30	
	%	N	%	N	%	N	%	N
Infinite	10.0		19.0		28.0		37.0	
10	0.0	0	0.0	0	0.0	0	0.0	0
20	0.0	0	0.0	0	0.0	0	2.4	1
30	0.0	0	0.0	0	2.5	1	8.7	3
50	0.1	1	2.8	2	8.2	5	15.1	8
100	3.0	3	7.5	8	14.0	14	21.5	22
200	5.1	11	10.9	22	18.1	37	26.1	53

Percentage Passing Subgroups Must Have (2-Year Improvement)

N	0		10		20		30	
	%	N	%	N	%	N	%	N
Infinite	19.0		27.1		35.2		43.3	
10	0.0	0	0.0	0	0.0	0	0.0	0
20	0.0	0	1.4	1	5.5	2	11.2	3
30	2.3	1	6.1	2	11.0	4	17.1	6
50	6.1	4	10.8	6	16.4	9	23.0	12
100	9.9	10	15.6	16	21.9	22	28.9	29
200	12.5	25	19.0	38	25.8	52	33.1	67

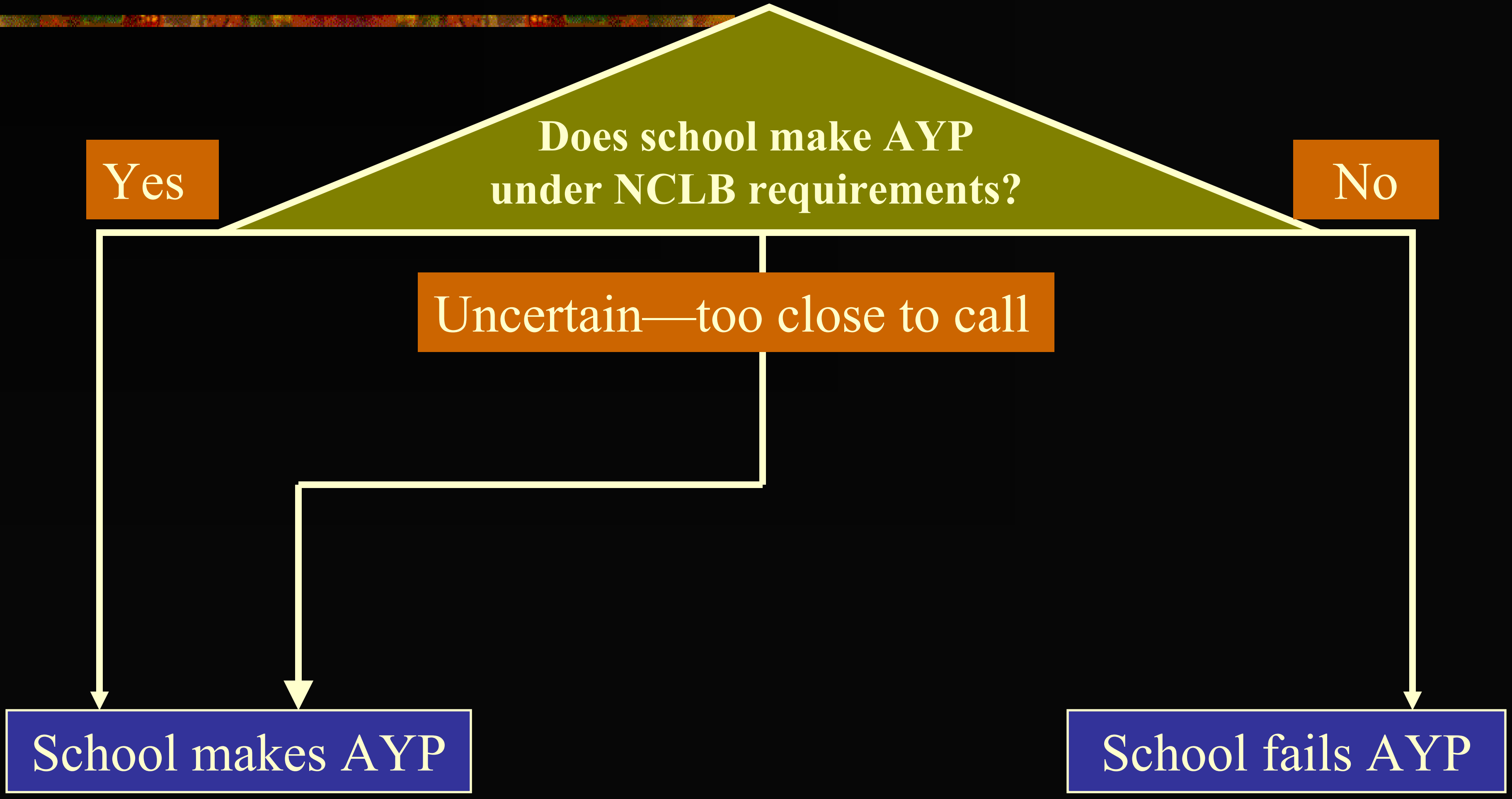
Percentage Passing Subgroups Must Have (3-Year Improvement)

N	0		10		20		30	
	%	N	%	N	%	N	%	N
Infinite	27.1		34.4		41.7		49.0	
10	0.0	0	0.0	0	0.0	0	0.0	0
20	4.0	1	5.2	2	8.7	2	13.7	3
30	8.2	3	10.5	4	14.7	5	20.2	7
50	12.5	7	15.9	8	20.8	11	26.7	14
100	16.8	17	21.3	22	26.9	27	33.2	34
200	19.8	40	25.1	51	31.2	63	37.8	76

Potential Solution 5: Take a Second Look

- Confidence intervals on status and growth
- When result is indeterminate (as it will be for most schools), use a second system of identification

One-Tiered System



Two (or more)-Tiered System